



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 117384**

**TO: Manjunath N Rao**  
**Location: REM/3B81/3C70**  
**Art Unit: 1652**  
**Monday, March 29, 2004**

**Case Serial Number: 09/847392**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**REM-1A55**  
**Phone: 571-272-2512**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Rao,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

117384

From: Rao, Manjunath N.  
Sent: Monday, March 22, 2004 10:57 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 09/847,392

From: Manjunath N. Rao  
Art Unit 1652, Room 3B81  
Mail Box in Room 3C70  
Phone: 272-0939

Date: 3-22-04

Please search the following as soon as possible for application with serial number

**09/847,392**

1. SEQ ID NO: 1 and nucleotides 557-1171 of SEQ ID NO:1 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2, against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao  
Art Unit 1652, Room 3B81,  
Mail Box in Room 3C70,  
Remsen Building, USPTO

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 3/24/04  
Date Completed: 3/24/04  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH: 2  
NA Sequences: \_\_\_\_\_  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 0371/05p  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

RECEIVED  
MAR 22 2004  
STIC

400<sup>th</sup> Dulany St.  
Alexandria, VA.  
Phone: 571-272-0939

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case.***





# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen E01 D86

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 28, 2004, 14:04:39 ; Search time 5020.17 Seconds  
(without alignments)  
10360.543 Million cell updates/sec  
Title: US-09-847-392-1  
Perfect score: 1200  
Sequence: 1 agaaataatgtgagatcgc.....atgtcggatcgcgcgtaaac 1200  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:  
2: gb\_htg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_or:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pl:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: em\_vi:  
30: em\_htg\_hum:  
31: em\_htg\_inv:  
32: em\_htg\_other:  
33: em\_htg\_mus:  
34: em\_htg\_pln:  
35: em\_htg\_rod:  
36: em\_htg\_mam:  
37: em\_htg\_vrt:  
38: em\_sy:  
39: em\_htgo\_hum:  
40: em\_htgo\_mus:  
41: em\_htgo\_other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1200	100.0	1200	6	AR172781	AR172781 Sequence
2	1200	100.0	1200	6	E41573	E41573 DNA encodin
3	1200	100.0	1231	6	E48927	E48927 Novel gene
4	1200	100.0	1231	6	AX030175	AX030175 Sequence
5	1189	99.1	91414	1	ECOW85	M87049 E. coli gen
6	1180	98.3	11509	1	AE000458	AE000458 Escherich
7	1159.4	96.6	290976	1	AE016990	AE016990 Shigella
8	1146.6	95.5	318703	1	AP002567	AP002567 Escherich
9	1128.2	94.0	301566	1	AE016769	AE016769 Escherich
10	754.6	62.9	258050	1	AL627278	AL627278 Salmonell
11	754.6	62.9	299991	1	AE016845	AE016845 Salmonell
12	753	62.7	21692	1	AE008884	AE008884 Salmonell
13	753	62.7	96086	1	STYSTMD1	AE233324 Salmonell
14	699	58.2	265383	2	AC020874	AC020874 Mus muscu
15	690	57.5	10253	1	AE015396	AE015396 Shigella
16	617	51.4	13301	1	AE005614	AE005614 Escherich
17	604.8	50.4	212936	2	AC020970	AC020970 Mus muscu
18	600.6	50.1	256373	2	AC020870	AC020870 Mus muscu
19	582.8	48.6	10592	1	AE005613	AE005613 Escherich
20	563	46.9	1576	1	ECPLDB	X0155 E. coli pld
21	527.8	44.0	9751	1	AE015397	AE015397 Shigella
22	417	34.8	417	6	AR352333	AR352333 Sequence
23	391.8	32.6	645	6	AR384166	AR384166 Sequence
24	361.4	30.1	10957	1	AE013640	AE013640 Yersinia
25	361.4	30.1	199050	1	AJ414159	AJ414159 Yersinia
26	297.2	24.8	1041	6	AR384122	AR384122 Sequence
27	294.4	24.5	214911	2	AC010537	AC010537 Homo sapi
28	270.4	22.5	268294	2	AC020885	AC020885 Mus muscu
29	243	20.2	214911	2	AC010537	AC010537 Homo sapi
30	233.2	19.4	206758	2	AC016129	AC016129 Drosophil
31	211.8	17.7	62274	2	AC020833	AC020833 Mus muscu
32	207.4	17.3	263	6	AR352166	AR352166 Sequence
33	191.4	16.0	63847	2	AC101459	AC101459 Mus muscu
34	165.4	13.8	11305	1	AE004104	AE004104 Vibrio ch
35	165.4	13.8	29938	1	AE016800	AE016800 Vibrio vu
36	162.6	13.5	245560	1	AP005330	AP005330 Vibrio vu
37	156	13.0	288108	1	AP005083	AP005083 Vibrio pa
38	146.8	12.2	3131	1	AB105408	AB105408 Photobact
39	144.8	12.1	1038	6	AR375232	AR375232 Sequence
40	138.4	11.5	110000	2	AC074221_2	Continuation (3 of
41	133	11.1	349907	1	EX571874	EX571874 Photorhab
42	112.4	9.4	10475	1	AE004937	AE004937 Pseudomon
43	107.6	9.0	313518	1	AE016856	AE016856 Pseudomon
44	91.4	7.6	11785	1	AE015905	AE015905 Shewanell
45	87.4	7.3	303226	1	AE016774	AE016774 Pseudomon

ALIGNMENTS

RESULT 1  
AR172781  
LOCUS AR172781 1200 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 1 from patent US 6303348.  
ACCESSION AR172781  
VERSION AR172781.1 GI:17912272  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1200)  
AUTHORS Livshits, V. Arkadievich., Zakataeva, N. Pavlovna.,  
Aleoshin, V. Varyamiovich., Balareova, A. Valentinovna. and  
Tokhmakova, I. Lvovna.  
TITLE DNA coding for protein which confers on bacterium escherichia coli

Pred. No. is the number of results predicted by chance to have a

resistance to L-homoserine and method for producing L-amino acids  
Patent: US 6303348-A 1 16-OCT-2001;  
Location/Qualifiers  
1. .1200  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 1 AGAAATATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
QY 61 GGACCGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCAATTAAG 120  
Db 61 GGACCGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCAATTAAG 120  
QY 121 CGATTACATGCCCCGAGATCGGATCGCTTAACAGCGACCGGACGTCCTCTGCCCGGA 180  
Db 121 CGATTACATGCCCCGAGATCGGATCGCTTAACAGCGACCGGACGTCCTCTGCCCGGA 180  
QY 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
QY 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTACACCCGCTCATGCTGTGTGC 300  
Db 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTACACCCGCTCATGCTGTGTGC 300  
QY 301 CGAAACGGACAAAGCGCACCGGATGTCATCCACACAGTAACTCTGTTTCATCACGC 360  
Db 301 CGAAACGGACAAAGCGCACCGGATGTCATCCACACAGTAACTCTGTTTCATCACGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCGTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCGTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTCTGTAACACATCGGTAATCTGCTCTTAAACACACGTAATCGT 480  
Db 421 TCCAGTCTTTTGTCTGTAACACATCGGTAATCTGCTCTTAAACACACGTAATCGT 480  
QY 481 TTTTGTAGCGTCTGACACAAACGCTGCGACAGTACGCTATTGTGGCACAATAATAGAC 540  
Db 481 TTTTGTAGCGTCTGACACAAACGCTGCGACAGTACGCTATTGTGGCACAATAATAGAC 540  
QY 541 ACACCGGGAGTTTATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGGAGTTTATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTGCGCAGGCTCTGTTGCAATCAACATATGACCACTCGCTCAACACCG 660  
Db 601 TTTAACGCTGTGCGCAGGCTCTGTTGCAATCAACATATGACCACTCGCTCAACACCG 660  
QY 661 TTATCCGGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATTCAATTTGTC 720  
Db 661 TTATCCGGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATTCAATTTGTC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCCGCTCAGTATGCGTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCCGCTCAGTATGCGTTTGAAGTTGAA 780  
QY 781 GTGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGTGGTGC 840  
Db 781 GTGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGTGGTGC 840  
QY 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGCGTGCACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGCGTGCACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGTATTTCCGCAATT 960

Db 901 TTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGTATTTCCGCAATT 960  
QY 961 CATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGGCGTCAACCACTATTGT 1020  
Db 961 CATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGGCGTCAACCACTATTGT 1020  
QY 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCCCTTCTCAACGGATTGCTCTATGGAT 1080  
Db 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCCCTTCTCAACGGATTGCTCTATGGAT 1080  
QY 1081 TAAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
QY 1141 GGGAGCGCTGTTAGCATCGCGGAGGCATCGTGAAATAATATGTCCGATCGGCGTAAAC 1200  
Db 1141 GGGAGCGCTGTTAGCATCGCGGAGGCATCGTGAAATAATATGTCCGATCGGCGTAAAC 1200

RESULT 2  
E41573  
LOCUS E41573 1200 bp DNA linear PAT 31-JAN-2002  
DEFINITION DNA encoding protein imparting tolerance against L-homoserine to Escherichia coli and process for producing L-amino acid.  
ACCESSION E41573  
VERSION E41573.1 GI:18627533  
KEYWORDS JP 2000116390-A/1.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 1200)  
AUTHORS Rivishittsu, V.A., Zakataeva, N.P., Aryoshin, V.V., Beraryo, A.V. and Tokumakova, I.R.  
TITLE DNA encoding protein imparting tolerance against L-homoserine to Escherichia coli and process for producing L-amino acid  
JOURNAL Patent: JP 2000116390-A 1 25-APR-2000;  
COMMENT AJINOMOTO CO INC  
OS Escherichia coli  
PN JP 2000116390-A/1  
PD 25-APR-2000  
PF 12-OCT-1999 JP 1999289777  
PR 13-OCT-1998 RU 98118425  
PI VITARI ARUKAJEVIHI RIVISHITTSU, NATARIYA PAVUROVUNA ZAKATAEVA,  
PI VLADIMIR VENIYAMINOVICHI ARYOSHIN, ARA VALENTINOVUNA BERARYO, PI IRINA RIVOVUNA TOKUMAKOVA  
PC C12N15/09, C07K14/245, C12N1/21, C12P13/06, C12P13/08// (C12N1/21, PC C12R1:19),  
PC C12N15/00  
CC  
FH Key Location/Qualifiers  
FT CDS (557)..(1171).  
FEATURES  
source Location/Qualifiers  
1. .1200  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 1 AGAAATAATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
QY 61 GGACCGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCAATTAAG 120  
Db 61 GGACCGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCAATTAAG 120  
QY 121 CGATTACATGCCCCGAGATCGGATCGCTTAACAGCGACCGGACGTCCTCTGCCCGGA 180  
Db 121 CGATTACATGCCCCGAGATCGGATCGCTTAACAGCGACCGGACGTCCTCTGCCCGGA 180  
QY 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
QY 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTACACCCGCTCATGCTGTGTGC 300  
Db 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTACACCCGCTCATGCTGTGTGC 300  
QY 301 CGAAACGGACAAAGCGCACCGGATGTCATCCACACAGTAACTCTGTTTCATCACGC 360  
Db 301 CGAAACGGACAAAGCGCACCGGATGTCATCCACACAGTAACTCTGTTTCATCACGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCGTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCGTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTCTGTAACACATCGGTAATCTGCTCTTAAACACACGTAATCGT 480  
Db 421 TCCAGTCTTTTGTCTGTAACACATCGGTAATCTGCTCTTAAACACACGTAATCGT 480  
QY 481 TTTTGTAGCGTCTGACACAAACGCTGCGACAGTACGCTATTGTGGCACAATAATAGAC 540  
Db 481 TTTTGTAGCGTCTGACACAAACGCTGCGACAGTACGCTATTGTGGCACAATAATAGAC 540  
QY 541 ACACCGGGAGTTTATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGGAGTTTATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTGCGCAGGCTCTGTTGCAATCAACATATGACCACTCGCTCAACACCG 660  
Db 601 TTTAACGCTGTGCGCAGGCTCTGTTGCAATCAACATATGACCACTCGCTCAACACCG 660  
QY 661 TTATCCGGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATTCAATTTGTC 720  
Db 661 TTATCCGGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATTCAATTTGTC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCCGCTCAGTATGCGTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCCGCTCAGTATGCGTTTGAAGTTGAA 780  
QY 781 GTGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGTGGTGC 840  
Db 781 GTGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGTGGTGC 840  
QY 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGCGTGCACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGCGTGCACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGTATTTCCGCAATT 960

QY 121 CGATTAAATGCCCCGAGATGGGATCGGCTAAACAGGGCGACCGGAACGTCCTGCCCGCGA 180  
Db 121 CGATTAAATGCCCCGAGATGGGATCGGCTAAACAGGGCGACCGGAACGTCCTGCCCGCGA 180  
QY 181 TGGTCGATGATTAAACATCAAAACCCCAAATGGAACAGGTCATAGGCCAGTTCGCGATAT 240  
Db 181 TGGTCGATGATTAAACATCAAAACCCCAAATGGAACAGGTCATAGGCCAGTTCGCGATAT 240  
QY 241 TTTACGTAGCTCTCAATACGCCCGCGGCAGATGACTACCACCGGTCTGCTGTGCG 300  
Db 241 TTTACGTAGCTCTCAATACGCCCGCGGCAGATGACTACCACCGGTCTGCTGTGCG 300  
QY 301 CGAAAACGGACAAAGCGCACCGGAATGTCTACACACAGTAAGAACTCTGCTTCATCAGC 360  
Db 301 CGAAAACGGACAAAGCGCACCGGAATGTCTACACACAGTAAGAACTCTGCTTCATCAGC 360  
QY 361 TGACGGCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGGTTTCTCTTGT 420  
Db 361 TGACGGCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGGTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTTAAACACCGTAAATCGT 480  
Db 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTTAAACACCGTAAATCGT 480  
QY 481 TTTTTTAGCGTGCCTGACACAAAGCTGCGACAGTAGCGTATTGTGGCAAAAATAGAC 540  
Db 481 TTTTTTAGCGTGCCTGACACAAAGCTGCGACAGTAGCGTATTGTGGCAAAAATAGAC 540  
QY 541 ACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTGCGCAGGCTCTGCTGCAATCAACACTATGACCACTCGCTCAACACGG 660  
Db 601 TTTAACGCTGTGCGCAGGCTCTGCTGCAATCAACACTATGACCACTCGCTCAACACGG 660  
QY 661 TTATCCGCGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATCATATTGTGC 720  
Db 661 TTATCCGCGCGGTGGCGTCTATTGCTGGCTTCAGACCGGACTGGCGATCATATTGTGC 720  
QY 721 TGGTTGGCGTGGGGTGGGACGCTATTTTCCCGCTCAGTGAATTGCGTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGGTGGGACGCTATTTTCCCGCTCAGTGAATTGCGTTGAAGTTGAA 780  
QY 781 GTGGGAGGCGCGGTTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC 840  
Db 781 GTGGGAGGCGCGGTTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC 840  
QY 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTGTTTCTGSGGGCGCTATTTCGGCAATT 960  
Db 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTGTTTCTGSGGGCGCTATTTCGGCAATT 960  
QY 961 CATCATGCCGCAACAGCGCAACTGATGCAGTATATCGTGTGCGGCTCACCACATTTGT 1020  
Db 961 CATCATGCCGCAACAGCGCAACTGATGCAGTATATCGTGTGCGGCTCACCACATTTGT 1020  
QY 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGAT 1080  
Db 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGAT 1080  
QY 1081 TAAAGGACCAAAAGCAGATGAAGGCGCTGAATAAGATTTTTCGCTCGTTGTTTATGCTGT 1140  
Db 1081 TAAAGGACCAAAAGCAGATGAAGGCGCTGAATAAGATTTTTCGCTCGTTGTTTATGCTGT 1140  
QY 1141 GGGAGCGCTGTTAGCATCGGCGAGGCATGCGTGAAAAAATATGTTCGGATCGGCGGTAAC 1200  
Db 1141 GGGAGCGCTGTTAGCATCGGCGAGGCATGCGTGAAAAAATATGTTCGGATCGGCGGTAAC 1200

RESULT 3  
E48927  
LOCUS 1231 bp DNA linear PAT 31-JAN-2002  
DEFINITION Novel gene and process for producing amino acid.  
ACCESSION E48927  
VERSION E48927.1 GI:18627996  
KEYWORDS JP 2000189177-A/1.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 1231)  
AUTHORS Rivishittsu,V.A., Zakataeva,N.P., Aryoshin,V.V., Beraryo,A.V. and Tokumakova,I.R.  
TITLE Novel gene and process for producing amino acid  
JOURNAL Patent: JP 2000189177-A 1 11-JUL-2000;  
AJINOMOTO CO INC  
COMMENT OS Escherichia coli  
PN JP 2000189177-A/1  
PD 11-JUL-2000  
PF 15-DEC-1999 JP 1999356018  
PR 23-DEC-1998 RU 98123511  
PI VITARI ARUKAJEVICHI RIVISHITTSU,NATARIYA PAVUROVUNA  
ZAKATAEVA,  
PI VLADIMIR VENIYAMINOVICHI ARYOSHIN,ARA VALENTINOVUNA BERARYO,  
PI IRINA RIVOVUNA TOKUMAKOVA  
PC C12N15/09,C07K14/245,C12N1/21,C12P13/06,C12P13/08,C12P13/08,  
PC C12P13/08//  
PC (C12N15/09,C12R1:19), (C12P13/06,C12R1:19), (C12P13/08,C12R1:19), PC  
C12N15/00,  
PC (C12N15/00,C12R1:19)  
CC  
FH Key Location/Qualifiers  
FT CDS (557)..(1171).  
FEATURES  
source  
1..1231  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
ORIGIN  
Query Match 100.0%; Score 1200; DB 6; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
QY 61 GGACCGGGCTGAACCTCTGCTGCCAGAATGCCGAGATCATCAACATAATCATTTAAAG 120  
Db 61 GGACCGGGCTGAACCTCTGCTGCCAGAATGCCGAGATCATCAACATAATCATTTAAAG 120  
QY 121 CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCGA 180  
Db 121 CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCGA 180  
QY 181 TGGTCGATGATTAAAGACATCAACCCCAAATGGAACAGGTCTATAGGCCAGTTCGCGCATAT 240  
Db 181 TGGTCGATGATTAAAGACATCAACCCCAAATGGAACAGGTCTATAGGCCAGTTCGCGCATAT 240  
QY 241 TTTACGTAGCTCTCAATACGCCCGCGGCAGATGACTACCACCGGTCTGCTGTGCG 300  
Db 241 TTTACGTAGCTCTCAATACGCCCGCGGCAGATGACTACCACCGGTCTGCTGTGCG 300  
QY 301 CGAAAACGGACAAAGCGCACCGGAATGTCTACACACAGTAAGAACTCTGCTTCATCAGC 360  
Db 301 CGAAAACGGACAAAGCGCACCGGAATGTCTACACACAGTAAGAACTCTGCTTCATCAGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGGTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGGTTTCTCTTGT 420



QY 421 TCCCAGTCTTTTGTGCTGTAACATCGGGTAATCTGCCTCTTAAACCAACGTAATAATCGT 480  
Db 421 TCCCAGTCTTTTGTGCTGTAACATCGGGTAATCTGCCTCTTAAACCAACGTAATAATCGT 480  
QY 481 TTTTATTTAGCGTGCCTGACACAAACGTCGACAGTAGCGTATTGTGGCACAAAATAGAC 540  
Db 481 TTTTATTTAGCGTGCCTGACACAAACGTCGACAGTAGCGTATTGTGGCACAAAATAGAC 540  
QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTCCCGAGGCTCTGGTGGCAATCAACACTATGACACACCTCGCTCAACCAACGG 660  
Db 601 TTTAACGCTGTCCCGAGGCTCTGGTGGCAATCAACACTATGACACACCTCGCTCAACCAACGG 660  
QY 661 TTATCCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATTGTGC 720  
Db 661 TTATCCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATTGTGC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTGGGTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTGGGTTTGAAGTTGAA 780  
QY 781 GTGGGACGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC 840  
Db 781 GTGGGACGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC 840  
QY 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATT 960  
Db 901 TTTTGTGAATCTCACCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATT 960  
QY 961 CATCATGCGCAACAGCGCAACTGATGACAGTATATCGTGCTCGCGCTCACCACTATTGT 1020  
Db 961 CATCATGCGCAACAGCGCAACTGATGACAGTATATCGTGCTCGCGCTCACCACTATTGT 1020  
QY 1021 GGTGATATTATTGTATGATCGGTAGCGCCACCCTTGCTCAACGGATTGCTCTATGGAT 1080  
Db 1021 GGTGATATTATTGTATGATCGGTAGCGCCACCCTTGCTCAACGGATTGCTCTATGGAT 1080  
QY 1081 TAAAGGACCAAGCAGATGAAGCGGTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAGCAGATGAAGCGGTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
QY 1141 GGGAGCGCTGTTAGCATCGCGGAGGATCGTGAAAAATAATGTCGGATGCGCGGTAAC 1200  
Db 1141 GGGAGCGCTGTTAGCATCGCGGAGGATCGTGAAAAATAATGTCGGATGCGCGGTAAC 1200

RESULT 4  
AX030175  
LOCUS AX030175 1231 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent EP1013765.  
ACCESSION AX030175  
VERSION AX030175.1 GI:10190392  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1  
AUTHORS Belareva,A.V., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L. and Zakataeva,N.P.  
TITLE Gene and method for producing l-amino acids  
JOURNAL Patent: EP 1013765-A 1 28-JUN-2000;  
AJINOMOTO KK (JP)  
FEATURES  
source location/Qualifiers  
1. .1231

/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"  
557. .1174  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC09318.1"  
/db\_xref="GI:10190393"  
/db\_xref="REMTREMBL:CAC09318"  
/translation="MTLEWFAAYLLTSLILTLSPGSGAINTMTTSLNHYGPAGGVYCV  
ASDRTGDSYAGWRGVGTLFSSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKLSLA  
STQSRRLHFORAVFVNLNPKSIVFLAALFPQFIMPPQPOLMQYIVLGVTTIVDIIV  
MIGYATLAQRTALWIKPKQKALNKIFGSLFVLVGLLASARHA"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
QY 61 GGACCGGCTGAACCTCCTGCTGCCAGAATGCCCGCAGATCATCAACATAATCATTAAG 120  
Db 61 GGACCGGCTGAACCTCCTGCTGCCAGAATGCCCGCAGATCATCAACATAATCATTAAG 120  
QY 121 CGATTAAATGCCCGAGATCGCGATCGGTAAACAGGACCGGAACGTCCTGCCCGCA 180  
Db 121 CGATTAAATGCCCGAGATCGCGATCGGTAAACAGGACCGGAACGTCCTGCCCGCA 180  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTAGGCCAGTTCGGCATAT 240  
Db 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTAGGCCAGTTCGGCATAT 240  
QY 241 TTTACGTAGTCTCAATACGCCCGGCGAGATGATACCAACCGGTCTGCTGTGCG 300  
Db 241 TTTACGTAGTCTCAATACGCCCGGCGAGATGATACCAACCGGTCTGCTGTGCG 300  
QY 301 CGAAAAACGACAAAGCGCACCGGAATGTCATCCACACAGTAACTCTGCTTCATCACGC 360  
Db 301 CGAAAAACGACAAAGCGCACCGGAATGTCATCCACACAGTAACTCTGCTTCATCACGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCTCTTAAACACCGTAAATCGT 480  
Db 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCTCTTAAACACCGTAAATCGT 480  
QY 481 TTTTATTTAGCGTGCCTGACACAAACGCTGCGACAGTAGCGTATTGTGGCACAAAATAGAC 540  
Db 481 TTTTATTTAGCGTGCCTGACACAAACGCTGCGACAGTAGCGTATTGTGGCACAAAATAGAC 540  
QY 541 ACACCGGGAGTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGGAGTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTGCCAGGCTCTGGTGAATCAACACTATGACACACCTCGCTCAACCAACGG 660  
Db 601 TTTAACGCTGTGCCAGGCTCTGGTGAATCAACACTATGACACACCTCGCTCAACCAACGG 660  
QY 661 TTATCCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATTGTGC 720  
Db 661 TTATCCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATTGTGC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTGGGTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTGGGTTTGAAGTTGAA 780  
QY 781 GTGGGACGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC 840



VNLFTRHRNNIRLTAAAGEKLLPYAETLMSTWQAARKRWRIPHDITSLVSPAPRCGNV  
CLISGDACIKIRMPIQIAYSSKRELENGSLW"  
complement (1136. .1165)  
/note="promoter matrix score of 63"  
1151. .1179  
/note="promoter matrix score of 57"  
1223. .1636  
/gene="o137"  
1223. .1636  
/gene="o137"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA67568.1"  
/db\_xref="GI:148172"  
/translation="MAESFTTNNRYFDNKHYPGRFSRHGDFTIKEAQLLERHGYAFNE  
LDLGKREPVTTEEKLFVAVCRGEREPVTEAERVWSKYMTRIKRPRFHTLSAVNRRLK  
VLKTTPIILINKGEMPPFPYVSSFRCRWSKRSIAR"  
1562. .1586  
/gene="o137"  
/note="terminator-like sequence"  
complement (1585. .3135)  
/gene="f516"  
complement (1585. .3135)  
/gene="f516"  
/function="unknown"  
/note="similar to Mg chelatase subunit of Rhodobacter  
capsulata"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA67569.1"  
/db\_xref="GI:148173"  
/translation="MHIMVTQGGMLSLSVHTRAALGVNAPPIITVEVHISKGLPGLTM  
VGLPETTVKARDVRSAIINSGYEYPAKKITINLAPADLPKEGGRYDLPIAIALAA  
SEQLTANKLDEVVLVGLALTGALRGVPGAISSATEPAIKSGRKIIIVAKDNEDEVGLIN  
GEGCLLIADHLQAVCAFLEGHKHALEXPRPTDAVSRAHQHDLSDVICQEQQKRGLEITAA  
GGHLLLIIGPGTGKTMASRINGLLPDLISNEEALESAAILSLVNAESVQKWRQRPF  
RSPHSAASLTAMVGGGAIIPGPGEISLAHNGVFLDELPEFERRITDALREPIESQI  
LSRTRAKITYPARFQLVAAMNPSPPTGHYQGNHNRCTPEQTLRYLNRLSGPFLDRFDLS  
LEIPLPPGILSKTVVPGESSATVKQVMAARERQFKQNKLNWLDSPEIRQFCKLE  
SEDAMWLEGTLIHLGLSIRAWQRLKLVARTIADIDQSDIITRQHLQEAVSYRAIDRL  
IHLQKLLT"  
misc\_structure 3287. .3433  
/note="predicted bend of 79 degrees"  
3297. .3326  
/note="promoter matrix score of 74"  
3336. .12791  
/note="corresponds to M10313 (1. .9456)"  
3336. .3737  
/note="corresponds to V00289 (1. .402)"  
3336. .3676  
/note="corresponds to M11651 (1. .341)"  
3379. .5721  
/note="V00290; ECILVX (1. .2343)"  
3458. .3556  
/gene="ilvL"  
3458. .3556  
/gene="ilvL"  
/codon\_start=1  
/transl\_table=11  
/product="ilvGMDA operon leader peptide"  
/protein\_id="AAA67570.1"  
/db\_xref="GI:148174"  
/translation="MTALLRVLISLVVISVVVVIIPPCGAALGRGKA"  
3696. .4679  
/gene="ilvG"  
3696. .4679  
/gene="ilvG"  
/note="CGSC No. 603; inactive in wild-type E. coli K12"  
/codon\_start=1  
/transl\_table=11  
/product="acetylhydroxy acid synthase II, large subunit"  
/protein\_id="AAA67571.1"

/db\_xref="GI:148175"  
/translation="MNGAQVVVHALRAQGVNTVFGYPGAIIMPVVDALYDGGVEHLIC  
RHEQGAAMAAGIYARATGKTGVCIAATSGPATNLITGLADALLDSIPVVAITQVSAP  
FIGTDAFQEVLDVGLSLACTKHSFLVQSLIEELPRIMABAFDVACSGRPGPVLVDIPKD  
IQLASGDLEPWFTEVNEVFPFHAEEVQAROMLAKPMLYVGGGVGMAQAVPALRE  
FLAATKMPATCTKLGLGAVEADYPYVLGMLGMHGTKAANFAVQECDDLIIAVGARFDDR  
VTGKLNTPFAPHASVIHMDIDPAEMNKLRQAHVALQGLNALLPALQQLNQ"  
4675. .5340  
/gene="O221"  
4675. .5340  
/gene="O221"  
Query Match 99.1%; Score 1189; DB 1; Length 91414;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
DB 62717 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 62658  
QY 61 GGACCGGGCTGAACCTCCTGCTGCCAGAATGCCGAGATCATCAACATAATCAATAAAG 120  
DB 62657 GGACCGGGCTGAACCTCCTGCTGCCAGAATGCCGAGATCATCAACATAATCAATAAAG 62598  
QY 121 CGATTAAACATGCCCGAGATCGGATCGGTAACAGGCGACCGGAACGTCCCTGCCCGCGA 180  
DB 62597 CGATTAAACATGCCCGAGATCGGATCGGTAACAGGCGACCGGAACGTCCCTGCCCGCGA 62538  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAAAATGGAACAGGTCTATAGGCCAGTTCCGCATAT 240  
DB 62537 TGGTCGATGATTAAGACATCAAAACCCCAAAATGGAACAGGTCTATAGGCCAGTTCCGCATAT 62478  
QY 241 TTTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCCCGGTCTGTTGCGG 300  
DB 62477 TTTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCCCGGTCTGTTGCGG 62418  
QY 301 CGAAAAACGACAAAGCGCACCGGAATGTCTCCACACAGTAACTCTGCTTCATCACGC 360  
DB 62417 CGAAAAACGACAAAGCGCACCGGAATGTCTCCACACAGTAACTCTGCTTCATCACGC 62358  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTTCTCTGTT 420  
DB 62357 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTTCTCTGTT 62298  
QY 421 TCCAGTCTTTTGTGCTGTAACATCGGGTAATCTGCCTCTTAAACCCAGTAAATCGT 480  
DB 62297 TCCAGTCTTTTGTGCTGTAACATCGGGTAATCTGCCTCTTAAACCCAGTAAATCGT 62238  
QY 481 TTTTATAGCGTCCCTGACACAAACGCTCGGACAGTAGCGTATTGTGGCACAAAAATAGAC 540  
DB 62237 TTTTATAGCGTCCCTGACACAAACGCTCGGACAGTAGCGTATTGTGGCACAAAAATAGAC 62178  
QY 541 ACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGTGACATCGATCAT 600  
DB 62177 ACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGTGACATCGATCAT 62118  
QY 601 TTAAACGCTGTGCCAGGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACCCACGG 660  
DB 62117 TTAAACGCTGTGCCAGGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACCCACGG 62058  
QY 661 TTATC-CGGCCGGTGGCGTCTATTGTGCGGCTTCAGACCGGACTGGGCATTTCATATTGTG 719  
DB 62057 TTATCSCGGCCGGTGGCGTCTATTGTGCGGCTTCAGACCGGACTGGGCATTTCATATTGTG 61998  
QY 720 CTGGTTGGCGGTGGGTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGA 779  
DB 61997 CTGGTTGGCGGTGGGTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGA 61938  
QY 780 AGTGGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGCGCGCCCGCTGGTG 839  
DB 61937 AGTGGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGCGCGCCCGCTGGTG 61878  
QY 840 CAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAG 899

Db 61877 CAATTGACCTTAATCGCTGGCCTCTACTCAATCGCGTCGACATTGTTCCAGCGCGCAG 61818  
QY 900 TTTTGTGAATCTCACCACATCCAAAAGTATTGTGTTCTGCGCGCGCTATTCCGCAAT 959  
Db 61817 TTTTGTGAATCTCACCACATCCAAAAGTATTGTGTTCTGCGCGCGCTATTCCGCAAT 61758  
QY 960 TCATCATGCCGCAACAGCCGCACTGATGCAGTATATCGTCTCGCGCTACCACTATTG 1019  
Db 61757 TCATCATGCCGCAACAGCCGCACTGATGCAGTATATCGTCTCGCGCTACCACTATTG 61698  
QY 1020 TGGTCGATATTATTGTCATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGA 1079  
Db 61697 TGGTCGATATTATTGTCATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGA 61638  
QY 1080 TTAAAGGACCAAGCAGATGAAGGCGCTGAATAGATTTTTCGGCTCGTTGTTATGCTGG 1139  
Db 61637 TTAAAGGACCAAGCAGATGAAGGCGCTGAATAGATTTTTCGGCTCGTTGTTATGCTGG 61578  
QY 1140 TGGGAGCGCTGTTAGCATCGGCGGAGGCGATCGTGAAAAAATAATGTCGGATGCGCGGTAAA 1199  
Db 61577 TGGGAGCGCTGTTAGCATCGGCGGAGGCGATCGTGAAAAAATAATGTCGGATGCGCGGTAAA 61518  
QY 1200 C 1200  
Db 61517 C 61517

RESULT 6  
AE000458/c  
LOCUS  
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete genome. linear BCT 01-DEC-2000  
ACCESSION AE000458 U00096  
VERSION AE000458.1 GI:2367299  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli K12  
Escherichia coli K12  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE  
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.  
The complete genome sequence of Escherichia coli K-12  
Science 277 (5331), 1453-1474 (1997)  
97426617  
9278503  
REFERENCE  
AUTHORS Blattner,F.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
REFERENCE  
AUTHORS Blattner,F.R.  
TITLE Direct Submission  
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459  
REFERENCE  
AUTHORS Plunkett,G. III.  
TITLE Direct Submission  
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
On Sep 9, 1997 this sequence version replaced gi:1790254.  
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director).  
Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently

FEATURES  
source

Location/Qualifiers  
1. 11509  
/organism="Escherichia coli K12"  
/mol\_type="genomic DNA"  
/strain="K12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:83333"  
81..950  
/gene="plda"  
/note="synonym: b3821"  
81..950  
/gene="plda"  
/EC number="3.1.1.32"  
/function="enzyme; Central intermediary metabolism: Phosphorus compounds"  
/note="O289; 99 pct identical to PA1\_ECOLI SW: P00631; CGSCNo. 384"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane phospholipase A"  
/protein\_id="AAC76824.1"  
/db\_xref="GI:2367300"  
/translation="MRTLQGLLPVFMPLMAYAQEATVKEVHDAPVGRSIIANMLQ EHDNPFTLYPYDTNYLIYQTSDLNKEAIAASYDWAENARKDEVKQSLAFPLWRGIL GENSVLGASYTQKSWQLNSSESPFRETNYEPQLFVGATDYRFAGTWLDRVEMGY NHDSNGRSDPTSRSNWRLYTRLMAENGNWLVVEKVPWVVGVDNDNDPDTIKYMGYYQLK IGYHLGDAVLGSAKQYNNWNTGYGGAELGSLYPITKHVRLYTQVYSGYGESLIDYNFNO TRVGVGMLNDLF"  
965..993  
/note="factor Sigma70; predicted +1 start at 4003392"  
1019..1047  
/note="factor Sigma70; predicted +1 start at 4003446"  
1077..2909  
/gene="recQ"  
/note="synonym: b3822"  
1077..2909  
/gene="recQ"  
/EC number="3.6.1.-"  
/function="enzyme; DNA - replication, repair, restriction/modification"  
/note="O610; 99 pct identical to 607 amino acids of RECQ\_ECOLI SW: P15043 but has 3 additional N-ter residues"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent DNA helicase"  
/protein\_id="AAC76825.1"  
/db\_xref="GI:2367301"  
/translation="MNVAQAEVLNLESGAKQVLOETFGYQQFRPGQEEIIDTVLSGRD CLVVMPTGGQKSLCYQIPALLNLGLTVVVSPLISLMKDQVDQLQANGVAAACLNSTQT

determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amherst.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.





Db 3989 GCAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCA 3930  
QY GTTTTTCGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCGCAA 958  
Db 3929 GTTTTTCGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCGCAA 3870  
QY TTCATCATGCGCGCAACAGCCGCAACTGATGCAGTATATCGTGTCTGGCGGTCCACCACTATT 1018  
Db 3869 TTCATCATGCGCGCAACAGCCGCAACTGATGCAGTATATCGTGTCTGGCGGTCCACCACTATT 3810  
QY GTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGG 1078  
Db 3809 GTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGG 3750  
QY ATTAAGACCAAAAGCAGATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTG 1138  
Db 3749 ATTAAGACCAAAAGCAGATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTG 3690  
QY GTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCCGATCGCGCGTAA 1198  
Db 3689 GTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCCGATCGCGCGTAA 3630  
QY 1199 AC 1200  
Db 3629 AC 3628  
RESULT 7  
AE016990  
LOCUS Shigella flexneri 2a str. 2457T section 13 of 16 of the complete genome. linear BCT 22-APR-2003  
DEFINITION  
ACCESSION AE016990 AE014073  
VERSION AE016990.1 GI:30042943  
KEYWORDS  
SOURCE  
ORGANISM Shigella flexneri 2a str. 2457T  
Shigella flexneri 2a str. 2457T  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.  
REFERENCE 1 (bases 1 to 290976)  
AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.  
Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T  
Infect. Immun. 71 (5), 2775-2786 (2003)  
12704152  
REFERENCE 2 (bases 1 to 290976)  
AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.  
Direct Submission  
Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
LOCATION/Qualifiers  
1. 290976  
/organism="Shigella flexneri 2a str. 2457T"  
/mol\_type="genomic DNA"  
/strain="2457T"  
/serotype="2a"  
/db\_xref="taxon:198215"  
complement(68. 1405)  
/gene="yjdE"  
/locus\_tag="S3622"  
complement(68. 1405)  
/gene="yjdE"  
/locus\_tag="S3622"  
/function="putative transport; Not classified"  
/note="residues 1 to 445 of 445 from Escherichia coli K-12 : residues 1 to 445 of 445 from Escherichia coli K-12 :  
gene  
CDS  
Db 3989 GCAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCA 3930  
QY GTTTTTCGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCGCAA 958  
Db 3929 GTTTTTCGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCGCAA 3870  
QY TTCATCATGCGCGCAACAGCCGCAACTGATGCAGTATATCGTGTCTGGCGGTCCACCACTATT 1018  
Db 3869 TTCATCATGCGCGCAACAGCCGCAACTGATGCAGTATATCGTGTCTGGCGGTCCACCACTATT 3810  
QY GTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGG 1078  
Db 3809 GTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGG 3750  
QY ATTAAGACCAAAAGCAGATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTG 1138  
Db 3749 ATTAAGACCAAAAGCAGATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTG 3690  
QY GTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCCGATCGCGCGTAA 1198  
Db 3689 GTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCCGATCGCGCGTAA 3630  
QY 1199 AC 1200  
Db 3629 AC 3628

B4115"  
/codon\_start=1  
/transl\_table=11  
/product="putative amino acid/amine transport protein"  
/protein\_id="AAP18666.1"  
/db\_xref="GI:30042944"  
/translation="MSSDADAHKVLIPVILMVSGNIMSGVFLLPANLASTGGIAIY  
GMLVTIIGALGLSMVYAKMSPFLDPSPGGSYAYARRCFGLGYQTNVYLWACWIGNI  
AMVIVGVLYSYFPPIKLDPLVLTITCVVVLWIFVLNLIVGPKMITRVOAVATVLALI  
PIVGIAVFGWFERGETYMAAWNVSGLGTGFGAIQSLTFLWSFIGVESASVAAGVVK  
NPKRNVPITATIGGVLIAAVCVLSTTAIMGMIIPNAALRVASPFGDAAARMALGDTAGA  
IVSFCAAAGCGLSGGWTLLAGCTAKAAADGDLFPPIFARVKNKAGTPVAGLIIVGILM  
TIFQLSSISPNATKEFGLVSSVSVIFTLVPYLYTCAALLLGHGHFGKARPAYLAVTT  
IAFLYCIWAVVSGAKEVMWVSFVTLMTVITAMYALNYNRLHKNPYPLDAPISKD"  
complement(1542. .2303)  
/gene="adiY"  
/locus\_tag="S3623"  
complement(1542. .2303)  
/gene="adiY"  
/locus\_tag="S3623"  
/function="putative regulator; Not classified"  
/note="residues 1 to 253 of 253 are 93.28 pct identical to  
residues 1 to 253 of 253 from Escherichia coli K-12 :  
B4116"  
/codon\_start=1  
/transl\_table=11  
/product="putative ARAC-type regulatory protein"  
/protein\_id="AAP18667.1"  
/db\_xref="GI:30042945"  
/translation="MRICSDPCIVLLTEKDVIRVANGKEPISLXANHMAALLNCENNI  
IDVSSLNNTLVAHISHDIIKDYLRFNLKDLSPVWQSRATPILTPCLTPDVFRVAA  
QHSMPAETESSEKERTALLFTVLSRFLDSKKFLSLMMYMLRNCVSDSVYQIESDIH  
KDWNLMSVASCLSLSPLLKKLKSNTSYSQITTCRMRYAVNELMMDGKNISQVSQ  
SCGYNSTSYFISVFKDFYGMTPLHYVSQHRRTVA"  
complement(2628. .4895)  
/gene="adia"  
/locus\_tag="S3624"  
complement(2628. .4895)  
/gene="adia"  
/locus\_tag="S3624"  
/function="enzyme; Degradation of small molecules: Amino  
acids".  
/note="residues 1 to 755 of 755 are 99.47 pct identical to  
residues 2 to 756 of 756 from Escherichia coli K-12 :  
B4117"  
/codon\_start=1  
/transl\_table=11  
/product="biodegradative arginine decarboxylase"  
/protein\_id="AAP18668.1"  
/db\_xref="GI:30042946"  
/translation="MKVLIVSEFLHQDTWVGNNAVERLADALSQQNVTVIKSTSFDDG  
FAILSSNEAIDCLMFQYQMEHPDEHQNVRLQIGLHERQONVPVFLGDRKALAAAD  
RDLELVDEFWILEDATDFIAGRAVAANTRYRQQLPLPLFSALMKYSDIHEYSWAAP  
GHQGVGFTKTTPAGRFYHDYGENLFRDTMGIERITSLGSLDHTGAFGESEKYYAARVF  
GADRSWVVTGSGSNRTIMQACMTDNDVVVDNRNCHKSTIEQGLMLTGAKPVYMWPSR  
NRYGIIGPIYPQEMQPETLQKKISESLPKDKAGQKPSYCVVNTCTYDGVYNAAKEAQ  
DLLEKTSDRLHDFDEAWYGYARFNPIYADHYAMRGEPEGDHNGPTVEATHSIHKLNLALS  
QASYIHVREGGAINFSRFNOAYMMHATSPLYAICANDVAVSMMDGNLSGLSLTQEV  
IDEAVDFRQMARLYKEFTADSGSWFEKPNKVVTDPTQTKTYDFAADPTKLLTTVQD  
CWVHPGESWHGFKDIPDNWSMLDPIKVSILAPMGDEGELEETGVPAALVTAWLGRH  
GIVPTRTTDFQIMFLFSMGVTRGKGLTVNTLCSFKHYVDANTPLAQVMPBELVEQYPD  
TYANMGIHDLGDTMFALWKENNPARGALNEAYSGLPVAEITPREAYNAIVDNNVELVSI  
ENLPGRIAAANSVIPYPPGIPMLLSGENFGDKNSPQSVYLSRLSQSWDHHFPRFEHETEG  
TEIIDGIYHVMCVKA"  
complement(5094. .6002)  
/gene="melR"  
/locus\_tag="S3625"  
complement(5094. .6002)  
/gene="melR"  
/locus\_tag="S3625"  
/function="regulator; Degradation of small molecules:  
Carbon compounds"  
/note="residues 1 to 302 of 302 are 99.66 pct identical to

residues 1 to 302 of 302 from Escherichia coli K-12 :

```
B4118"
/codon_start=1
/transl_table=11
/product="regulator of melibiose operon"
/protein_id="AAP18669.1"
/db_xref="GI:30042947"
/translation="MNTDPMFCSSDEKOTRSPLSLYSEYQRMIEFRAPHIMPTSHWH
QOEVNVPDGDVEYLINNEKVINQGHITLFWACPHQLTDTGTCQSMALFNLPMHL
FLSWPLDKDLINHVTGHMVKSLATQQLSPFEVRRWQOELNSPNEIROQLAIDIGLM
LKRFLSSWEPILVNKTSTHNSVSRHAQFVYSQMLGFIAENYDQALTINDVAEHVK
LNANYAMGIFORVMQLTMKQYITAMRINHVRALLSDTKSILDIALTAGFRSSSRFYS
TFGKVVGMSPQYRKLSQRRQTFFG"
6060. .6114
/note="BOXC-20"
6285. .7640
/gene="meIA"
/locus_tag="S3626"
6285. .7640
/gene="meIA"
/locus_tag="S3626"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="residues 1 to 451 of 451 are 98.89 pct identical to
residues 1 to 451 from Escherichia coli K-12 :
B4119"
/codon_start=1
/transl_table=11
/product="alpha-galactosidase"
/protein_id="AAP18670.1"
/db_xref="GI:30042948"
/translation="MMSAPKITFIFAGSTIFVKNILGDVFHREALKTAHIALMDIDPT
RLEESHIVVRKLMDSAGSGKITCTHQKEALEDADFVVAFOIGYEPCTVTDPEVC
KRHGLEQTIADTLPGGGINRALRTIPHLWQICENMTEVCPDATMLNVPMAWNTWAM
YARYPIKQVGLCHSVQGTABELARDLINIDPATLRYRCAGINHMAYLELERKTADGS
VYNLYPELLAAEAGQAPKPNHNGNTRCNIVRYEMPKLGYFVTESSEHFAEYTPWF
IKLGREDLIERKVPLEDPKRCVEQLANWHKELEYKNASRIDIKPSREYASTIMNA
INTGEPSTVIVNVRNDGLIDNLPQGCCVEVACLVDANGIOPTKVGTLPShLAALMQTN
INVQTLTTEALTENRDRVYHAAMDPHPTAAVLGIDBIYALVDLIAAHGMDLPSWLH
R"
7755. .9164
/gene="meIB"
/locus_tag="S3627"
7755. .9164
/gene="meIB"
/locus_tag="S3627"
/function="transport; Transport of small molecules:
Carbohydrates, organic acids, alcohols"
/note="residues 1 to 469 of 469 are 96.37 pct identical to
residues 1 to 469 from Escherichia coli K-12 :
B4120"
/codon_start=1
/transl_table=11
/product="melibiose permease II"
/protein_id="AAP18671.1"
/db_xref="GI:30042949"
/translation="MTTKLSYGFAGFGKDPFAGIVVMYLMYYTVDVGLSVGLVGTLP
LVARIWDAINDPIMGWIVNATRSRWGKFPWILIGTLANSVILFLFSAHLFEGTQOI
VFVCVTYILWMTYTIMDIPFWSLVPTITLDKRERELVYPYPRFFASLAGFTVAGVTL
PFVNVYGGDRGFQMFMTLVLIATFFIVSTIITLRNVHEVFSNQSSAEGSHLTKA
IVALIYKNDQISCLLGMALAYNASNIITGFAYIYFSYVIGDADLPYIYLSYAGANL
VTLVFFRLVKLSRRILWAGASILPVLSCGVLLMALMSYHNVLIVIAIGILLNVGT
ALFWLVQIVMADTVDYGEYKLVHRCESIAYSQTMVVKGSFAFAFFIAVLGMIGY
VPNVEQSTQALLGMQFIMIALPTLFFVMVTLILYFRFRLNGDITLRIQIHLLDKYRKV
PPEPVHADIPYGAUSDVKA"
9250. .9283
/note="REP (repetitive extragenic palindromic) element,
contains 1 REP sequence"
complement(9303. .9932)
/gene="yjdf"
/locus_tag="S3628"
complement(9303. .9932)
/gene="yjdf"
```

```
/locus_tag="S3628"
/note="residues 1 to 209 of 209 are 83.25 pct identical to
residues 1 to 209 from Escherichia coli K-12 :
B4121"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP18672.1"
/db_xref="GI:30042950"
/translation="MTRTLKPLILNTGALALTLLIYTGISAHDKLTWLEVTPIIV
VPLLLATARRVPLTLLYTLIFPHAILMVGGQYTYAKVPVFEVQEWLGLSRNPYDK

Query Match      96.6%; Score 1159.4; DB 1; Length 290976;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1189; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY      1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db      251314 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 25131373

QY      61 GGACCGGGCTGAACCTCCTGCTGCCAGNATGCCGCCAGATCATCAACATAATCAATAAAG 120
Db      251374 GGACCGGGCTGAACCTCCTGCTGCCAGNATGCCGCCAGATCATCAACATAATCAATAAAG 251433

QY      121 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGGGACCGGAACGTCCTCCCTGCCCGCGA 180
Db      251434 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGGGACCGGAACGTCCTCCCTGCCCGCGA 251493

QY      181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCCGCATAT 240
Db      251494 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCCGCATAT 251553

QY      241 TTTACGTAGCTCTCAATAGCCCGGGCAGATGACTACCAACCCGGTCATGGTGTCTGTSCG 300
Db      251554 TTTACGTAGCTCTCAATAGCCCGGGCAGATGACTACCAACCCGGTCATGGTGTCTGTSCG 251613

QY      301 CGAAAAACGGACAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCATCAGCG 360
Db      251614 CGAAAAACGGACAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCATCAGCG 251673

QY      361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAACAGCAGCAACCGGTTTTCTCTTGT 420
Db      251674 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAACAGCAGCAACCGGTTTTCTCTTGT 251733

QY      421 TCCAGTCTTTTGTGCTGTAACATCGGGTAACTCGCTCTTTAAACCCACGTAAATCGT 480
Db      251734 TCCAGTCTTTTGTGCTGTAACATCGGGTAACTCGCTCTTTAAACCCACGTAAATCGT 251793

QY      481 TTTTATTAGCGTGCCTGACACAAACGCTCGGACAGTAGCGTATTGTGGCACAAAAATAGAC 540
Db      251794 TTTTATTAGCGTGCCTGACACAAACGCTCGGACAGTAGCGTATTGTGGCACAAAAATAGAC 251853

QY      541 ACACCGGGAGTTTCATCATGACCTTAGATGGTGGTTGGCTTACCTGCTGACATCGATCAT 600
Db      251854 ACACCGGGAGTTTCATCATGACCTTAGATGGTGGTTGGCTTACCTGCTGACATCGATCAT 251913

QY      601 TTAAACGCTGTCGCCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGG 660
Db      251914 TTAAACGCTGTCGCCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGG 251973

QY      661 TTAT--CCGSCCGGTGGCTCTATTGCTGGGCTTCAGACCCGGACTGCGGATTCATATTGT 718
Db      251974 TTATCGCGCGCGGTGGCTCTATTGCTGGGCTTCAGACCCGGACTGCGGATTCATATTGT 252033

QY      719 GCTGGTTGGCGTGGGGTT--GGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGT 777
Db      252034 GCTGGTTGGCGTGGGGTTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGT 252093

QY      778 GAAGTGGGCAGGCGCGGCTTACTTTGATTGGTGGGAATCCAGCAGTGGCGCGCGCTGG 837
Db      252094 AAAGTGGGCAGGCGCGGCTTACTTTGATTGGTGGGAATCCAGCAGTGGCGCGCGCTGG 252153

QY      838 TGCAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTGTTGCCAGCGCGC 897
```

||||| 252154 CGCAATTGATCTTAATCGCTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGC 252213

||||| 898 AGTTTTTGTGAATCTCACCAAATCCCAAAGTATTGTGTTCTCGCGCGGCTATTTCGGCA 957

||||| 252214 AGTTTTTGTGAATCTCACCAAATCCCAAAGTATTGTGTTCTCGCGCGGCTATTTCGGCA 252273

||||| 958 ATTCAATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGCGCTCACCACCTAT 1017

||||| 252274 ATTCAATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGCGCTCACCACCTAT 252333

||||| 1018 TGTGTCGATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATG 1077

||||| 252334 TGTGTCGATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATG 252393

||||| 1078 GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCT 1137

||||| 252394 GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCT 252453

||||| 1138 GGTGGAGCGCTGTTAGCATCGCGAGGCGATCGTGAAAAATAATGTCGGATCGCGCGTA 1197

||||| 252454 GGTGGAGCGCTGTTAGCATCGCGAGGCGATCGTGAAAAATAATGTCGGATCGCGCGTA 252513

||||| 1198 AAC 1200

||||| 252514 AAC 252516

RESULT 8

AP002567/c 318703 bp DNA linear BCT 07-MAR-2001

LOCUS Escherichia coli O157:H7 DNA, complete genome, section 18/20.

DEFINITION AP002567 BA000007

ACCESSION AP002567.1 GI:13364198

VERSION

KEYWORDS

SOURCE Escherichia coli O157:H7

ORGANISM Escherichia coli O157:H7

REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

AUTHORS 1 (sites)

Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

Genes Genet. Syst. 74 (5), 227-239 (1999)

JOURNAL 20198780

MEDLINE 10734605

PUBMED

REFERENCE 2 (sites)

AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.

TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MGL655

Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

JOURNAL 20557356

MEDLINE 11108008

PUBMED

REFERENCE 3 (sites)

AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.

TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak

Gene 258 (1-2), 127-139 (2000)

JOURNAL 20564182

MEDLINE 11111050

PUBMED

REFERENCE 4 (sites)

AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,

TITLE Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.

JOURNAL Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12

MEDLINE DNA Res. 8 (1), 11-22 (2001)

PUBMED 21156231

REFERENCE 11258796

AUTHORS 5 (bases 1 to 318703)

Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)

COMMENT genome project.

FEATURES

source

1. .318703

/organism="Escherichia coli O157:H7"

/mol\_type="genomic DNA"

/strain="O157:H7"

/sub\_strain="RIMD 0509952"

/db\_xref="taxon:83334"

103. .648

/gene="ECs4723"

103. .648

/gene="ECs4723"

/note="similar to WECD ECOLI gi|1790225 percent identity 98 in 181 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB38146.1"

/db\_xref="GI:13364199"

/translation="MQAKJAASNTGELDALQQLGFSVEGEVDLALPVNNASDSGAVV AQETDIPALRQLASAAFAQSRFRAPWYAPDASSRFYAQWIENAVRGTFDQCILIRAA SGDIRGYVSLRELNATDARIGLLAGRGAGAEMLQALTALNWAYARGKTTLRVATQMGNAT ALKRYIQSGANVESTAYWLYR"

553. .1783

/gene="ECs4724"

553. .1783

/gene="ECs4724"

/note="similar to WECE ECOLI gi|2367285 percent identity 99 in 376 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="putative regulator"

/protein\_id="BAB38147.1"

/db\_xref="GI:13364200"

/translation="MIPFNAPPVVGTELDMQSAMSGKLCGDGGFTRRCQWLEORF GSAKVLTPSCTASLEMAALLDIQGDDEVIMPSYTFVSTANAFVLRGAKIVFDVRP DTMNIDETLIEAAITDKTRVIVPVHYAGVACEMDTMALAKHNLNFVEDAAQGVNST YKGRALGTICHIGCFSFHETKXNYTAGEGGATLINDKALIERAEIIREKGTNRSQFPR GQVDKYTWRDIGSSYLSMDLQAAYLWALQLEAADRINQORLALWQNYDALAPLAKAGR IELPSIPDCQLQNAHMFYIKLRDIDGRSALINFLKRAEIMAVFHYIPLHGCPAGERFG EFHGEDRYTTKESERLLRLPLFYNLSPVQRTVIATLLNYFS"

1785. .3035

/gene="ECs4725"

1785. .3035

/gene="ECs4725"

/note="similar to WZXE ECOLI gi|1790227 percent identity 98 in 416 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="putative cytochrome"

/protein\_id="BAB38148.1"

/db\_xref="GI:13364201"





Db 37186 CGCAATTGATCTTAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGGC 37127  
QY 898 AGTTTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCA 957  
Db 37126 AGTTTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCA 37067  
QY 958 ATTATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGGCGTCACCACTAT 1017  
Db 37066 ATTATCATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGGTCGTCACCACTAT 37007  
QY 1018 TGTGTCGATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATG 1077  
Db 37006 TGTGTCGATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATG 36947  
QY 1078 GATTAAGGACCAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCT 1137  
Db 36946 GATTAAGGACCAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCT 36887  
QY 1138 GGTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCGGATCGGCGTA 1197  
Db 36886 GGTGGAGCGCTGTTAGCATCGCGAGGCATCGGTGAAAAATAATGTCGGATCGGCGTA 36827  
QY 1198 AAC 1200  
Db 36826 AAC 36824

RESULT 9  
AE016769/c  
LOCUS  
DEFINITION Escherichia coli CFT073 section 15 of 18 of the complete genome.  
ACCESSION AE016769 AE014075  
VERSION AE016769.1 GI:26110701  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli CFT073  
Escherichia coli CFT073  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 301566)  
AUTHORS Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P., Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J., Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C., Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.  
Extensive Mosaic Structure Revealed by the Complete Genome Sequence of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
12471157  
2 (bases 1 to 301566)  
AUTHORS Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P., Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J., Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C., Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.  
Direct Submission  
Submitted (20-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES  
source  
1. .301566  
/organism="Escherichia coli CFT073"  
/mol\_type="Genomic DNA"  
/strain="CFT073"  
/db\_xref="taxon:199310"  
complement(72. .1100)  
/gene="waaW"  
/locus\_tag="c4450"  
complement(72. .1100)  
/gene="waaW"  
/locus\_tag="c4450"  
/note="Residues 1 to 341 of 342 are 98.24 pct identical to residues 1 to 341 from Genpept.129 :  
>gb|AAC69673.1| (AF019746) UDP-galactose: (galactosyl) LPS  
alpha1,2-galactosyltransferase WaaW [Escherichia coli]"  
/codon\_start=1

/transl\_table=11  
/product="UDP-galactose: (galactosyl) LPS  
alpha1,2-galactosyltransferase"  
/protein\_id="AA082886.1"  
/db\_xref="GI:26110702"  
/translation="MDLLAESITEVAVSGEIANTRVNLNIAIGIDRNFLFGAAVSMQS  
VVMHNPDLAVKEHFLTDTYIDEDYLQRVNAFTSKNANVEVRIYKVSSAFIDIPSLKQW  
SYATFFRLVAFQYLSSETIENLLYIDADVICKGSLAGLLDINFDGDKFAAVIKDVPFMQ  
EXPAKRLAIEGLPGNYFNAGVYVQLQLEAWKNDFMKAIAMLASDPQKTKYKCLDQDI  
LNILFFSHCIFISGDYDCFYGIDYELKNKSDYKTIIDDTKLIHYVGVTKPWNDWT  
NYPCKYFNEAYQVSCWNVAFIPATNEKQYQVKYQHAKNGDITFNAFIYFIKPKLNK  
YXRKLFGQ"  
complement(1126. .1818)  
/gene="rfaj"  
/locus\_tag="c4451"  
complement(1126. .1818)  
/gene="rfaj"  
/locus\_tag="c4451"  
/function="enzyme; Macromolecule metabolism:  
Lipopolysaccharide"  
/note="Escherichia coli K-12 ortholog: b3625; Escherichia  
coli O157:H7 ortholog: z5052"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide core biosynthesis protein  
rfaj"  
/protein\_id="AA082887.1"  
/db\_xref="GI:26110703"  
/translation="MITSIRYRGFSFYKDNNDNKYEIPDEILAYNFKTVKVLNRNID  
TKVSLIDTKYGRYVFKVAPKTRNERFLKSFVKGDYQNLIVETDVRVSAGLTDPND  
FYPLAERKIFNYASVFIMLIEYVEGVELNDMPPIPEDVKAEIKASMEKHLALNMLSGD  
PHRGNFIVSKDGVRIIDLSGKSCSTAERKARDRLAMERHLGIANEIKDYGYYSVIYRTK  
LRKFIKLLKGKA"  
complement(1828. .2823)  
/gene="rfaj"  
/locus\_tag="c4452"  
complement(1828. .2823)  
/gene="rfaj"  
/locus\_tag="c4452"  
/EC\_number="2.4.1.58"  
/function="enzyme; Surface polysaccharides and antigens"  
/note="Escherichia coli K-12 ortholog: b3626; Escherichia  
coli O157:H7 ortholog: z5051"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide 1,2-glucosyltransferase"  
/protein\_id="AA082888.1"  
/db\_xref="GI:26110704"  
/translation="MNEFIKERFSYLADNKNENVPENLVSYGIDKNFLYGAGVISISY  
LINNSDINFVHFVFTDYDDDLKSFNETAKQNTSIIIVLIDPKYPADLPTSQFWSY  
ATYFRVLSFEYLSIESITLLYLDADVCKGSLKPLTKIIFKDEFAAVIPDNSTQAAC  
AKRLNIPGMNGRYFNAGVIYVNLKKWHEANLTPYLLTLRGETKYGSLKYLDQDALNI  
AFNMNNIYLAKDFDTIYTLKNELHDSHRKYYQQTITDKTVLIHYTGITKFWHSWAGYP  
SASYFNIAREQSPWKYPKLEARTVAEMQKQYKHLFAHGEYIKGITSIIKYKLKK"  
complement(2840. .3856)  
/gene="rfaj"  
/locus\_tag="c4453"  
complement(2840. .3856)  
/gene="rfaj"  
/locus\_tag="c4453"  
/EC\_number="2.4.1.44"  
/function="enzyme; Surface polysaccharides and antigens"  
/note="Escherichia coli K-12 ortholog: b3627; Escherichia  
coli O157:H7 ortholog: z5053"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide 1,3-galactosyltransferase"  
/protein\_id="AA082889.1"  
/db\_xref="GI:26110705"  
/translation="MSAHYFNPMQEMINKTIIFDERPAVSVASSFHVAYGIDQNFPGC  
GVSTSVLHNNDVSFVHFVIDDIPADIQRLSQAKSYHTCIQIHLVNCERLKALP  
TTKNWSIAMYFRFIADYDQDKILYLDADACQGNLKPILITMDLANNVAAVYTER  
DANWWSLRAQSLQCNELEKGYFNSGVLLINTLAWAQESVSAMSLADKAVVSRITY



Db	280910	CGCAATTGATCTTAATCGCTGGCTCCACTCAATCGCGTGAGACATTTGTTCCAGCGGC	280851
QY	898	AGTTTGTGTAATCTCACCAATCCCAAAAGTATGTGTTCTCGCGCGGCTATTTCGGCA	957
Db	280850	AGTTTGTGTAATCTCACCAATCCCAAAAGTATGTGTTCTCGCGCGGCTATTTCGGCA	280791
QY	958	ATTGATCATGCCGCAACAGCCGCAACTGATGCAGPATATCGTGCTCGGCGTCAACCACTAT	1017
Db	280790	ATTGATCATGCCGCAACAGCCGCAACTGATGCAGPATATCGTGCTCGGCGTCAACCACTAT	280731
QY	1018	TGTGTCGATATTATTGTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATG	1077
Db	280730	TGTGTCGATATTATTGTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATG	280671
QY	1078	GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTTATGCT	1137
Db	280670	GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTTATGCT	280611
QY	1138	GGTGGAGCGCTGTTAGCATCGCGGAGCGATCGGTGAAAAATAATGTCGGATCGGCGGTA	1197
Db	280610	GGTGGAGCGCTGTTAGCATCGCGGAGCGATCGGTG-AAATAATGTCGGATCGGCGGTA	280552
QY	1198	AAC 1200	
Db	280551	AAC 280549	
RESULT 10	AL627278	258050 bp	DNA linear BCT 04-JUL-2003
LOCUS	Salmonella enterica serovar Typhi (salmonella typhi) strain CT18,		
DEFINITION	complete chromosome, segment 14/20.		
ACCESSION	AL627278	AL513382	
VERSION	AL627278.1	GI:16504263	
KEYWORDS	Salmonella enterica subsp. enterica serovar Typhi		
SOURCE	Salmonella enterica subsp. enterica serovar Typhi		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
REFERENCE	1 (bases 1 to 258050)		
AUTHORS	Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Connor,P., Cronin,A., Davis,P., Davies,R.M., Dowd,B., White,N., Farrar,J., Feltwell,T., Hamlin,N., Hague,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.		
TITLE	Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18		
JOURNAL	Nature 413 (6858), 848-852 (2001)		
MEDLINE	21534947		
PUBMED	11677608		
REFERENCE	2 (bases 1 to 258050)		
AUTHORS	Parkhill,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	E-mail: parkhill@sanger.ac.uk		
Notes:	Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.		
FEATURES	(URL, http://www.sanger.ac.uk/Projects/S_typhi/).		
source	Location/Qualifiers		
	1. 258050		
	/organism="Salmonella enterica subsp. enterica serovar Typhi"		
	/mol_type="genomic DNA"		
	/strain="CT18"		
	/db_xref="taxon:90370"		
gene	69. .1544		

CDS	/gene="tolC"		
	/note="synonym: STY3364"		
	69. .1544		
	/gene="tolC"		
	/note="Similar to Salmonella enteritidis outer membrane protein TolC precursor tolC SW:TOLC_SALEN (Q54001) (491 aa) fasta scores: E(): 0, 99.0% id in 491 aa, and to Escherichia coli outer membrane protein tolC precursor tolC or mtcb or muka or refI SW:TOLC_ECOLI (P02930) (495 aa) fasta scores: E(): 0, 89.7% id in 495 aa Orthologue of E. coli tolC (TOLC_ECOLI); Fasta hit to TOLC_ECOLI (495 aa), 90% identity in 495 aa overlap"		
	/codon_start=1		
	/transl_table=11		
	/product="outer membrane protein TolC precursor"		
	/protein_id="CAD07712.1"		
	/db_xref="GI:16504264"		
	/db_xref="GOA:Q8Z3N8"		
	/db_xref="SPTREMBL:Q8Z3N8"		
	/translation="MQMKLLPILIGLSLGSFSTLSQAEMLMQVYQQARLSNPELRKS AADRDAAFEKINEARSPLLPOLGLGADYTSNGYRDANGINSNETSASLQTLTQDFM SKWRGLTLQEKAGIQDVTYQTDQOTLLINTANAYFKVLNAIDVLSYTAQKEAIYRQ LDQTTQRFNVGLVAITDVQNAQAQYDTVLANEVTARNLNDNAVEELRQVTGNYYPELA SLNVEHFKTDKPKAVNALLKEAENRNLSLQARLSQDLAREQIRQAQDGHLPITNLTA STGISDTSYSGSKTNSAQYDDSNMGQNKILNFSPLYQGMVNSQVKAQYNFVGAS EQLESARSVVQTVRSFNINASSINAYKQAVSAQSSLDAMEAGSVGTRTIIVD VLDATTTLYDAKQQLANARYTYLINQLNIKALGTNEQDLLALNSTLKGPIPTSPES VAPETPEQDAADGYNHAAPAVQPTAARANSNNGNPFRRH"		
gene	1757. .2428		
	/gene="STY3365"		
CDS	1757. .2428		
	/gene="STY3365"		
	/note="Orthologue of E. coli ygiB (YGIB_ECOLI); Fasta hit to YGIB_ECOLI (234 aa), 97% identity in 223 aa overlap. Contains a possible N-terminal signal sequence."		
	/codon_start=1		
	/transl_table=11		
	/product="possible lipoprotein"		
	/protein_id="CAD07713.1"		
	/db_xref="GI:16504265"		
	/db_xref="SPTREMBL:Q8XGZ1"		
	/translation="MKRTKSIHASFRRKSWSAARHLTPVALAVTAVFVLAGECKSDETV SLYQNADDCSAANPGKSAECTAYNNALKEAERTAPKYATREDCTVAEFGGCGCQQA QAGMAPENQAQOQSSGSFWMPLMAGYMGRLMGAGFAQQPLFSSKNPASPAYGKY TDAAGKNYGAAQPGRTMTVPKTAMAPKPATTTVTRGGFESVAKQSTMQRSAAGTST RSMGG"		
misc_feature	1835. .1867		
	/gene="STY3365"		
	/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"		
gene	2434. .3597		
CDS	2434. .3597		
	/gene="STY3366"		
	/gene="STY3366"		
	/note="Fasta hit to YJFC_ECOLI (387 aa), 50% identity in 392 aa overlap		
	Orthologue of E. coli ygiC (YGIC_ECOLI); Fasta hit to YGIC_ECOLI (386 aa), 94% identity in 386 aa overlap"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved hypothetical protein"		
	/protein_id="CAD07714.1"		
	/db_xref="GI:16504266"		
	/db_xref="SPTREMBL:Q8XFP6"		
	/translation="MERVSITERPDWRDKATEYGFNFHTWYGEYPWCEDAYVYKLTLAQ VEKLEDVTAELHQMLKVVVERVIASDELMTKFRIPKHTWGFVRQSWQTQPSLYSRLD LAMDGIGEPKLENNADTPTSLYEAAFFQWILEDDQINAGNLPEGSDQFNSLQEKLE RFALREQYGFQLLHLTCRTVEDRGTIQYLDQCAAEAEIATEFLYIDDIGLGEKGQ FTDLDQDVIANFLKLYPWEFLREMFSFKLEDAGVRWLEPAWKSIISNKALLPLLWEM FPDHPNLLPAYFAEDEHPMDKIVVKPIFSREGANVSIIENGKTIESVEGPGYEGEMI VQQFPLPKFGDSYTLIGSWLNDQAPAGIGREDRALITQDLSRFYHFIVEG"		
gene	complement (3659. .4489)		
	/gene="STY3367"		



```
CDS
complement(3659. .4489)
/ gene="STY3367"
/ note="Orthologue of E. coli ygiD (YGID_ECOLI); Fasta hit
to YGID_ECOLI (271 aa), 87% identity in 266 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="CAD07715.1"
/ db_xref="GI:16504267"
/ db_xref="GOA:Q8Z3N7"
/ db_xref="SPTREMBL:Q8Z3N7"
/ translation="MVMLTKSLCSCKNIMSLTCTMPALFLGHGSPMNVLDNDYTRAWR
RLGEALPRQAIIVVSAHWYTRGTGVTAMERPOTLHDEGGFPQALYDMHYAPGSPAL
AQRLLVELLAPVPVALDKEAWGFDHGSWGLIKMYFNADIPMVQLSVSDSTKPAWHFEM
GRKLATLRDEGVMLVASGNVHNLRTRVWHGDNIPYPAASFNDPVKANLTWQGPVEQ
HPLVNYLQHEGGALSNPTPEHFLPLLYVLGAWDGKEPITIPVDGIEMGSISLSLVQVG
"
4588. .5361
/ gene="STY3368"
4588. .5361
/ gene="STY3368"
/ note="Orthologue of E. coli ygiE (YGIE_ECOLI); Fasta hit
to YGIE_ECOLI (257 aa), 93% identity in 257 aa overlap.
Contains multiple possible membrane spanning hydrophobic
domains"
/ codon_start=1
/ transl_table=11
/ product="putative membrane protein"
/ protein_id="CAD07716.1"
/ db_xref="GI:16504268"
/ db_xref="GOA:Q8XGR4"
/ db_xref="SWISS-PROT:Q8XGR4"
/ translation="MSVPLILTLLAGAATPIGAFVLGVQKPSNRVLAFSLGFAAGIM
LLISLMEPLPALDTEGMSPLVGYGMFLIGLLGYFGLDRLLPHAHQDLVQKQOPLP
GSIKRTAILTLGISLNFPEGIATFTVASSNLELFGIALVALHNIPEGLAVAGPV
YAATGSKRTAIFWAGISGNWAEILGVVLAWLILGSLVSPIVMAAIAVAGIMVALSVD
ELMPLAKEIDPNNPNSYGVLCGMSIMGLSLVILQTIGIG"
6004. .7800
/ gene="STY3370"
/ note="synonym: asst"
6004. .7800
/ gene="STY3370"
/ note="Similar to Klebsiella pneumoniae arylsulfate
sulfotransferase asst TR:P97036 (EMBL:U32616) (598 aa)
fasta scores: E(): 0, 86.8% id in 598 aa"
/ codon_start=1
/ transl_table=11
/ product="probable arylsulfate sulfotransferase"
/ protein_id="CAD07717.1"
/ db_xref="GI:16504269"
/ db_xref="GOA:Q8Z3N6"
/ db_xref="SPTREMBL:Q8Z3N6"
/ translation="MFDQYRKTLILAGAVALTGGLTAASTFAAGFQPAQPAKLGAVVV
DPYGNAPLTALVELDSHIIISDVKTVHKGKGVFVTVYVTKESLETYDGIPIFLYQ
KFANNVTVEYKENGKAMKDDVYVQTSAINVHMDNRSISDLQOTKVIKVPAGFEDRLY
LVNTHFTFPQAEFWHGEKDNAGILDAGPAGGALPFDIAFYTFVVDVTQGEYRWLID
QDTFYNGHDMNINKRGYLMGIRETPRGFTFVQGOHWYFEDMMGOILADHKLPRGLD
ASHESIETVNGVLLRVGKRDYRKEDGIHVHTIRDQIIIEVDKSGRVVDVWDLTKILD
MRDALLGALDAGAVCVNVDLAHQQAQKLEPDTPYGDALGVGAGRNWAWHNSIAYDAK
DDSIILSSRHQGIKIGRDXQVKNILAPSKGNKQLASKLLKPVDDHGKPLTCDENGK
CKDTPDFDTYQHTAWLSSKGTITVFDNGDGRGLEQPALPTMKYSREVEYKIDKKGT
VQVWEYCKERGVDYFSPITSVVEYQXDRDTMFGFGGSINLFDVGKPTVGLKNEIDYK
TKEVKVEIDVLSKPNQTHYRALLVHPTQMFK"
7820. .8491
/ gene="dsba"
/ note="synonym: STY3371"
7820. .8491
/ gene="dsba"
/ note="Similar to Enterobacter amnigenus disulfide
isomerase dsba TR:Q9XDP1 (EMBL:AF012826) (222 aa) fasta
scores: E(): 0, 90.1% id in 222 aa"
/ codon_start=1
/ transl_table=11
```

```
misc_feature
7847. .8488
Query Match 62.9%; Score 754.6; DB 1; Length 258050;
Best Local Similarity 78.6%; Pred. No. 1.7e-223;
Matches 927; Conservative 0; Mismatches 249; Indels 3; Gaps 2;
QY 1 AGAAATAATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 247188 AGGAACAGGTTGGCTATCGCGCCCCCATTTAGTGGCCAGGATATAACGCTTACGCCAC 247247
QY 61 GGACCGGGTGAACCTCCTGCTGCCAGAAATGCCGACAGATCATCAACATAATCATTAAAG 120
Db 247248 GACCCCGGCTCAATTTCCCTGCTGCCAGAACGCCGCTAAGTCCTCAACATAATCGTTAAA 247307
QY 121 CGATTAAACATGCCCGAGATCGGGATCGGTAACAGCGGACCGGAACGTCCTGCCCGCGGA 180
Db 247308 TGATCGACATGACCCCGATCGGGTCCGATAACATCGCCCGGAGGCCCTTGACCACGG 247367
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAAAATGGAACAGAGTCATAGGCCAGTTCCCGCATAT 240
Db 247368 TGGTCGATGATGAAGATATCAAAACCCGAGATGAAAAGGTCATACGCCAGTTCCAGCGTAT 247427
QY 241 TTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCGGTCATGGTCTGTGG 300
Db 247428 TTACATAACTCTCAATGCGGCCAGGACAAATGACGATCGTCCGATCGTTGGAGTCATT 247487
QY 301 CGAAAACGGACAAAGCGCACCGGAATGTGCATCCACACAGTAGTAACCTGCTTTCATCACGC 360
Db 247488 CGAAAACGGACAAAGCGCACCGGAATGTGCATCCACCGCGGATAAACTCCGCTCTTCCCGC 247547
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 420
Db 247548 TGACGCCAAATAATCGGTGACGGGCCCATAGCAAAAGCGGCAAAACGCTTCTCTTGT 247607
QY 421 TCCAGTCTTTTGTGCTGTAAGAACATCGGGTAATCTGCTCTTTAAACCAACGTAATAATCGT 480
Db 247608 TCCAGTCAATTTGCTGCTGTAAGAACATCGGGTTTTCGGCTCGTTTACCAGGTAATAATCGT 247667
QY 481 TTTTTTTAGCGTCTGACAAACGCTGGGACAGTAGGCTATTGTGGCACAAAAAATAGAC 540
Db 247668 TTTTTCGTCGTGCTGACAAACACAGCGGCGAGTAGGCTATTGTGGCACAAAAAACAGAC 247727
QY 541 ACACCGGGAGTTCATCATGACCTTAGAATGGTGGTTTGGCTACCTGCTGACATCGATCAT 600
Db 247728 AACCAAGGAGTTTCGATGACCTTTGAATGGTGGTTTCGCCCTATCTACTGACCTCAACTCT 247787
QY 601 TTTAACGCTGTCGCCAGGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACACGG 660
Db 247788 GTTCACTTTCTCCGGGTTGAGCGGCCCATCAATACCATGACGACGCTATCAACCATGG 247847
QY 661 TTATCCGGCCG--GTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTCTATTGT 718
Db 247848 ATATCGTGGCGACGGCTTCTATCGCCGGGCTCCAGACCGGGCTGGGATACATATCGT 247907
QY 719 GCTGGTTGGCGGGGTTGGG--ACGCTATTTTCCCGCTCAGTGATTGGCTTTGAAGTGT 777
Db 247908 ACTGGTGGCGGTGCGGACTGGGTACGCTCTTTTCGCGCTCGCTCATCGCTTTTGAATCT 247967
QY 778 GAAGTGGCGAGCGCGCTTACTTGTATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
Db 247968 GAAATGGGCTGGCGCGCTTATCTTATCTGGCTGGGTATCCAGCAATGCGCGCGCGCAGG 248027
QY 838 TGCATTTGACCTTAAATCGCTGGCCTCTACTCAATCGCGCTCGACATTTGTTCCAGCGGC 897
```

Db 248028 CGCTATCGATCTGCATACTCTCGCCACAGCGCAATCGCGGTCGGCTGTTCAAACGGGC 248087

Qy 898 AGTTTTGTGAATCTCACCAATCCAAAAAGTATTGTCTTCTGGCGGCTATTTCCGCA 957

Db 248088 GATATTGTCAATCTAACCAATCCAAAAAGTATTGTCTTCTGCGCCCTGTTCCGCA 248147

Qy 958 ATTATCATGCCGCAACAGCCGCAACTGATGCAATATATCGTCTCGGCTCACCACATAT 1017

Db 248148 ATTATCATGCCGCAACAGCCGCAACTGCGCAGTACCTCAATCTCGGCTCACCACGAT 248207

Qy 1018 TGTGTTCGATATTATTGTGATGATCGGTTACGCCACCTTGTCAACGGATTGCTCTATG 1077

Db 248208 TGTGTTCGATATTATTGTGATGATCGGTTACGCCACCTTGTCAACGGATTGCGGCTG 248267

Qy 1078 GATTAAAGGACCAACAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCT 1137

Db 248268 GATTAAAGGACCAACAGCAGATGAAGCGCTGAATAAGCGTTGTTGTTGTTATGCT 248327

Qy 1138 GGTGGAGCGCTGTTAGCATCGGCGAGGCGATCGGTGAAA 1176

Db 248328 GGTAGCGCGCTCTCGGCTCGGCAAGACACCGCGTGATA 248366

RESULT 11

AE016845

LOCUS

DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 12 of 16 of the complete genome.

ACCESSION AE016845

VERSION AE016845.1

KEYWORDS GI:29139182

SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2

ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2

REFERENCE 1 (bases 1 to 299991)

AUTHORS Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G. F., Rose, D. J., Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.

TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18

JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)

MEDLINE 22531367

PUBMED 12644504

REFERENCE 2 (bases 1 to 299991)

AUTHORS Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J., Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.

TITLE Direct Submission

JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1. .299991

/organism="Salmonella enterica subsp. enterica serovar Typhi Ty2"

/mol\_type="genomic DNA"

/strain="Ty2"

/serovar="Typhi"

/sub\_species="enterica"

/db\_xref="taxon:205261"

complement(83. .709)

/gene="ftsJ"

/locus\_tag="t3214"

complement(83. .709)

/gene="ftsJ"

/locus\_tag="t3214"

/note="corresponds to STY3475 from Accession AL513382: Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="cell division protein"

/protein\_id="AAO70750.1"

/db\_xref="GI:29139183"

/translation="MTGKRSASSSRWLQEHFSDKYVQQAQKGLSRWFKLDEIQQ

SDKLFKPGMTVVDLGAAPGWSQYVVTQIGGKGRIIACDILLPMDPIVGVDFFLOQDFRD

ELVMKALLERVGDSKVQVYVMSDMPNMSGTPAVDI PRAMYLVELALDWC RDVLAPGGS

FVKVFGEGFDEYLRERSLFTKVKVRKPDSSRSRSEYIVATGRK"

797. .1129

/locus\_tag="t3215"

797. .1129

/locus\_tag="t3215"

/note="corresponds to STY3476 from Accession AL513382: Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAO70751.1"

/db\_xref="GI:29139184"

/translation="MTRFQSRKQKQYTMNLSTKQKHKLGLAHLKPKVVM LGNNGLTE

GVLAIEQALEHHEHLIKVXIASEDRETKTLIVDAIVRETGACNVQVICKTLVLYRPTK

BRKISLPR"

complement(1285. .1761)

/gene="greA"

/locus\_tag="t3216"

complement(1285. .1761)

/gene="greA"

/locus\_tag="t3216"

/note="corresponds to STY3477 from Accession AL513382: Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="transcription elongation factor"

/protein\_id="AAO70752.1"

/db\_xref="GI:29139185"

/translation="MQAIPMTLRGAELRBEELDFLKSVRPEIIAIAIAEAREHGD LKE

NAEYHAAREQCFCEGRIDIEAKLSAQVIDVTKMPNNGRVIFGATVTVLNLDTDEE

QYRIVGDDAEFKQNLISVNSPIARGLIGKEQDDVVVTKTPGGDVEEVLKVEYL"

2009. .3442

/locus\_tag="t3217"

2009. .3442

/locus\_tag="t3217"

/note="corresponds to STY3479 from Accession AL513382: Salmonella typhi CT18; D-alanyl-D-alanine carboxypeptidase"

/codon\_start=1

/transl\_table=11

/product="penicillin-binding protein"

/protein\_id="AAO70753.1"

/db\_xref="GI:29139186"

/translation="MRFSRFIIGLTTSIAFSVQAANIDBIYKQLPAGANLALMVQKIG

APAPADYHSQOMALPASTOKVITATAALIQLPDPFRFTTLETGKNVDNGILKGDLI

ARFGDPTLRQDIRNMVATLKSGVTQIDGNVLDTISIFASHDKAPGFWNDLTQCF

SAPAAAI VDRNCFSVSLVSAQKPNDLAFIRVASVYPTMFSQVRTLPRGSADAQYCE

LDVVPDNLRYLTGCLPQADPLPLAFAIQDGASAGAILKQELKEACITYRGTLRL

QTVNEPGTIVASKQSAPLHLLKIMLKSDNMIADTVFMIGHRVFNVPGTWRAGSD

AVRQLRQAGIDIGNTIADGSGLSRHNLIAPATMQVLQYIAQHDNLFISMLPL

AGYDGSLOYRAGLHQAGVDGKVSAGTSGLSQGVYNLAGFITTASGQRMAFVQYLSGYAV

PPADQNRRIPLVRFESRLYKDIYQNN"

complement(3573. .4745)

/locus\_tag="t3218"

complement(3573. .4745)

/locus\_tag="t3218"

/note="corresponds to STY3480 from Accession AL513382: Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="probable GTP-binding protein"

/protein\_id="AAO70754.1"

/db\_xref="GI:29139187"

/translation="MKFVDEASILVAGDGGNGCVSFRREKVI PKGGPDGGDGGDGD

VNMEADENLTLDYRFEKSPRAERQNGASRDCTKRGKDVTIKVPVGTREVIDQGTG

ETWGMTKHGQRLLVAKGGWHGLGNTRFKSSVNRTPRQKNTGTPGDKRDLLLEMLLA

DVGMGLMPNAGKSTFIRAVSAKPKVADYFTTLVPSLGVVRMDSEKSFVVADIPGLI

EGAAEGAGLIRFLKHLERCVRLLHLIDIDIDGSDPVENARIIGELEKYSQDLAAK

PRWLVFNKIDLMKSEAEKAKIAEALGWEGKYLIISAASQLGVKDLCDWDMTFIIE

NPIAQAEAAKQPEKVEFMWDDYHRQLAEVEEDADDDDDWDDDEDEEGVEFIYKR"

complement(4761. .5726)

```

CDS
/locus_tag="t3219"
/complement(4761..5726)
/locus_tag="t3219"
/note="corresponds to STY3481 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAO70755.1"
/db_xref="GI:29139188"
/translation="MKQAGIGILLALTAMCWGALPIAMKQVLEVNPEPSTIVFYRPL
MASIGLGAIALVAKRKLPLRIFRKPRWLVLAIATCGLGNFILFSSSLQYLSPASQ
VIGQLSPVGMVASVFILKERMGRTQVIGALMLLSGLVMFPNTSLIETRLTDYTWQ
VIFGGAAMVWSYGVAQKVLRLASQQILFLYTLCTIALLPLAKPMVIAQLSDWQ
LACLIIFGLNTLVGYGALAEAMARWQAQVSAILITLPLFTLLFSDLLSMAWPDFEAR
PMLNLGLGAFVYVAGAMYSAIGHRIWGLRKHETVVSQPRSGE"
/complement(5856..6113)
/gene="rpmA"
/locus_tag="t3220"
/complement(5856..6113)
/gene="rpmA"
/locus_tag="t3220"
/note="corresponds to STY3482 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L27"
/protein_id="AAO70756.1"
/db_xref="GI:29139189"
/translation="MAHKAGGSTRNGRDSIAKRLGVKFGGEAVLAGSIIVRQRTK
FHAGTNVCGGRDHLFAKADGVKFEVKGPKNRKYISIVAE"
/complement(6133..6444)
/gene="rplU"
/locus_tag="t3221"
/complement(6133..6444)
/gene="rplU"
/locus_tag="t3221"
/note="corresponds to STY3483 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L21"
/protein_id="AAO70757.1"
/db_xref="GI:29139190"
/translation="MYAVFQSGGKQHRVSEGTVRLEKLDIATGETIEFAEVLMIANG
BEVKIGVPFVDGGVKAIEVVAHGRGEXKIVKFRRRKHVRKQQGHRQWFDTVKITGIS
A"
/locus_tag="t3222"
/complement(6133..6444)
/gene="ispB"
/locus_tag="t3222"
/note="corresponds to STY3484 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="octaprenyl-diphosphate synthase"
/protein_id="AAO70758.1"
/db_xref="GI:29139191"
/translation="MNLKINELTAQDMAGVNVNATILEQLNSDVQLINOLGYVYIISGGG
KRIRPMIAVLAAVGYQGNHVTIAALIEFIHTATLLHDDVDVDESMDRKGATANAA
FGNAASVLYGDFIYTRAFQMTSLGSLKVLVMSVAVNVAEGLVQLMNVNDPDITE
ENYMRVYKTLARLFEAAACQSGILAGCTPEQEKGLQDYGRYLGTAFQLIDLLDYS
DGEHLGNVGDLDLNEGKPTLPLIHAMRHGTPEQSAMIRTAIEQNGRHLLEPVLNVT
TCGSLEWTRQRAEEADKAISALQLPDTPTWREALIGLAHIAVQDR"
/locus_tag="t3223"
/complement(6133..6444)
/gene="nlp"
/locus_tag="t3223"
/note="corresponds to STY3485 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="octaprenyl-diphosphate synthase"
/protein_id="AAO70759.1"
/db_xref="GI:29139192"
/translation="MESKLDIWHPADIIAGLEKKGTSMASRRRNLSSSTLANALTR
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="Ner-like regulatory protein"
/protein_id="AAO70759.1"
/db_xref="GI:29139192"
/translation="MESKLDIWHPADIIAGLEKKGTSMASRRRNLSSSTLANALTR
Query Match 62.9%; Score 754.6; DB 1; Length 299991;
Best Local Similarity 78.6%; Pred. No. 1.8e-223;
Matches 927; Conservative 0; Mismatches 249; Indels 3; Gaps 2;
Qy 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 132183 AGGAACAGGTTGGCTATCGCGCCCCCATTTGAGTGGCCAGGATATAACGCTTACGCCAC 132242
Qy 61 GGACCGGCTGAACCTCTGTGCCAGAAATGCCGCCAGATCATCAACATAATCATTAAG 120
Db 132243 GACCCGGCTCAATTTCTGTGCCAGAACCGCGCTAAGTCCTCAACATAATCGTTAAA 132302
Qy 121 CGATTAACATGCCCGAGATGCGGATCGGCTAACAGCGACCGGAAACGTCCTGCCCGGA 180
Db 132303 TGATCGACATGACCCCGATGCGGTCGCGATAAATCCGCCGAGCGCCCTTGACCACGG 132362
Qy 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGATAT 240
Db 132363 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGATAT 132422
Qy 241 TTTAGCTAGCTCTCAATACGCCCGGGCAGATGACTACACCCGGTCATCGGTGCTGTGG 300
Db 132423 TTTACATAACTCTCAATGCGGCAGGACAAATGACGATCGTCCGATCGTTGGAGTCATTC 132482
Qy 301 CGAAACCGGACAAAGCGCACCGGAAATGTATCCACACGCGGATAAACTCCGGCTCTTCCCGC 360
Db 132483 CGAAACCGGACAAAGCGCACCGGAAATGTATCCACACGCGGATAAACTCCGGCTCTTCCCGC 132542
Qy 361 TGACCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 420
Db 132543 TGACCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 132602
Qy 421 TCCAGTCTTTTCTGCTGAACATCGGGTAATCTGCCTCTTTAAACCCAGTAAATCGT 480
Db 132603 TCCAGTCTTTTCTGCTGAACATCGGGTAATCTGCCTCTTTAAACCCAGTAAATCGT 132662
Qy 481 TTTTITTAGCGTGTGACACACGCTCGGACAGTAGCGGATTTGTGGCACAAAAATAGAC 540
Db 132663 TTTTITTAGCGTGTGACACACGCTCGGACAGTAGCGGATTTGTGGCACAAAAATAGAC 132722
Qy 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600
Db 132723 AACCAAGGAGTTTTCGATGACCTTTGAATGGTGGTTTGCCTATCTACTGACCTCAACTCT 132782
Qy 601 TTTAACGCTGTGCGCAGGCTCTGCTGCAATCAACACTATGACCACTCGCTCGCTCAACCCACGG 660
Db 132783 GTTGAAGTCTTTCTCGGCTTCAGGCGCCATCAATACCATGACGACGTCATCAACCCATGG 132842
Qy 661 TTATCCGGCCG--GTGGCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGT 718
Db 132843 ATATCGTGGCGCAGCGGCTTCTATCGCGGGCTCCAGACCGGCTGGGGATACATATCGT 132902
Qy 719 GCTGTTGGCGTGGGTTGGG-ACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGT 777
Db 132903 ACTGTTGGCGTGGGTTGGG-ACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGT 132962
Qy 778 GAAGTGGCGAGGCGGCTTACTTGAATTTGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
Db 132963 GAAATGGGCTGGCGCGGCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCAGG 133022
Qy 838 TGCAATGACCTTAATCGTGGCTCTACTCAATCGCGTGCAGCATTTTGTCCAGCGCGC 897
Db 133023 CGCTATCGATCTGCATACTCTCGCCAGACGCAATCGCGGGGTGGCTGTTCACACGGGC 133082
Qy 898 AGTTTGTGTAATCTACCAATCCCAAGTATTGTGTTTCTGGCGCGCTATTTCGCCGA 957
```

```

CDS
/locus_tag="t3219"
/complement(4761..5726)
/locus_tag="t3219"
/note="corresponds to STY3481 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAO70755.1"
/db_xref="GI:29139188"
/translation="MKQAGIGILLALTAMCWGALPIAMKQVLEVNPEPSTIVFYRPL
MASIGLGAIALVAKRKLPLRIFRKPRWLVLAIATCGLGNFILFSSSLQYLSPASQ
VIGQLSPVGMVASVFILKERMGRTQVIGALMLLSGLVMFPNTSLIETRLTDYTWQ
VIFGGAAMVWSYGVAQKVLRLASQQILFLYTLCTIALLPLAKPMVIAQLSDWQ
LACLIIFGLNTLVGYGALAEAMARWQAQVSAILITLPLFTLLFSDLLSMAWPDFEAR
PMLNLGLGAFVYVAGAMYSAIGHRIWGLRKHETVVSQPRSGE"
/complement(5856..6113)
/gene="rpmA"
/locus_tag="t3220"
/complement(5856..6113)
/gene="rpmA"
/locus_tag="t3220"
/note="corresponds to STY3482 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L27"
/protein_id="AAO70756.1"
/db_xref="GI:29139189"
/translation="MAHKAGGSTRNGRDSIAKRLGVKFGGEAVLAGSIIVRQRTK
FHAGTNVCGGRDHLFAKADGVKFEVKGPKNRKYISIVAE"
/complement(6133..6444)
/gene="rplU"
/locus_tag="t3221"
/complement(6133..6444)
/gene="rplU"
/locus_tag="t3221"
/note="corresponds to STY3483 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L21"
/protein_id="AAO70757.1"
/db_xref="GI:29139190"
/translation="MYAVFQSGGKQHRVSEGTVRLEKLDIATGETIEFAEVLMIANG
BEVKIGVPFVDGGVKAIEVVAHGRGEXKIVKFRRRKHVRKQQGHRQWFDTVKITGIS
A"
/locus_tag="t3222"
/complement(6133..6444)
/gene="ispB"
/locus_tag="t3222"
/note="corresponds to STY3484 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="octaprenyl-diphosphate synthase"
/protein_id="AAO70758.1"
/db_xref="GI:29139191"
/translation="MNLKINELTAQDMAGVNVNATILEQLNSDVQLINOLGYVYIISGGG
KRIRPMIAVLAAVGYQGNHVTIAALIEFIHTATLLHDDVDVDESMDRKGATANAA
FGNAASVLYGDFIYTRAFQMTSLGSLKVLVMSVAVNVAEGLVQLMNVNDPDITE
ENYMRVYKTLARLFEAAACQSGILAGCTPEQEKGLQDYGRYLGTAFQLIDLLDYS
DGEHLGNVGDLDLNEGKPTLPLIHAMRHGTPEQSAMIRTAIEQNGRHLLEPVLNVT
TCGSLEWTRQRAEEADKAISALQLPDTPTWREALIGLAHIAVQDR"
/locus_tag="t3223"
/complement(6133..6444)
/gene="nlp"
/locus_tag="t3223"
/note="corresponds to STY3485 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="octaprenyl-diphosphate synthase"
/protein_id="AAO70759.1"
/db_xref="GI:29139192"
/translation="MESKLDIWHPADIIAGLEKKGTSMASRRRNLSSSTLANALTR
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="Ner-like regulatory protein"
/protein_id="AAO70759.1"
/db_xref="GI:29139192"
/translation="MESKLDIWHPADIIAGLEKKGTSMASRRRNLSSSTLANALTR
Query Match 62.9%; Score 754.6; DB 1; Length 299991;
Best Local Similarity 78.6%; Pred. No. 1.8e-223;
Matches 927; Conservative 0; Mismatches 249; Indels 3; Gaps 2;
Qy 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 132183 AGGAACAGGTTGGCTATCGCGCCCCCATTTGAGTGGCCAGGATATAACGCTTACGCCAC 132242
Qy 61 GGACCGGCTGAACCTCTGTGCCAGAAATGCCGCCAGATCATCAACATAATCATTAAG 120
Db 132243 GACCCGGCTCAATTTCTGTGCCAGAACCGCGCTAAGTCCTCAACATAATCGTTAAA 132302
Qy 121 CGATTAACATGCCCGAGATGCGGATCGGCTAACAGCGACCGGAAACGTCCTGCCCGGA 180
Db 132303 TGATCGACATGACCCCGATGCGGTCGCGATAAATCCGCCGAGCGCCCTTGACCACGG 132362
Qy 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGATAT 240
Db 132363 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGATAT 132422
Qy 241 TTTAGCTAGCTCTCAATACGCCCGGGCAGATGACTACACCCGGTCATCGGTGCTGTGG 300
Db 132423 TTTACATAACTCTCAATGCGGCAGGACAAATGACGATCGTCCGATCGTTGGAGTCATTC 132482
Qy 301 CGAAACCGGACAAAGCGCACCGGAAATGTATCCACACGCGGATAAACTCCGGCTCTTCCCGC 360
Db 132483 CGAAACCGGACAAAGCGCACCGGAAATGTATCCACACGCGGATAAACTCCGGCTCTTCCCGC 132542
Qy 361 TGACCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 420
Db 132543 TGACCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 132602
Qy 421 TCCAGTCTTTTCTGCTGAACATCGGGTAATCTGCCTCTTTAAACCCAGTAAATCGT 480
Db 132603 TCCAGTCTTTTCTGCTGAACATCGGGTAATCTGCCTCTTTAAACCCAGTAAATCGT 132662
Qy 481 TTTTITTAGCGTGTGACACACGCTCGGACAGTAGCGGATTTGTGGCACAAAAATAGAC 540
Db 132663 TTTTITTAGCGTGTGACACACGCTCGGACAGTAGCGGATTTGTGGCACAAAAATAGAC 132722
Qy 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600
Db 132723 AACCAAGGAGTTTTCGATGACCTTTGAATGGTGGTTTGCCTATCTACTGACCTCAACTCT 132782
Qy 601 TTTAACGCTGTGCGCAGGCTCTGCTGCAATCAACACTATGACCACTCGCTCGCTCAACCCACGG 660
Db 132783 GTTGAAGTCTTTCTCGGCTTCAGGCGCCATCAATACCATGACGACGTCATCAACCCATGG 132842
Qy 661 TTATCCGGCCG--GTGGCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGT 718
Db 132843 ATATCGTGGCGCAGCGGCTTCTATCGCGGGCTCCAGACCGGCTGGGGATACATATCGT 132902
Qy 719 GCTGTTGGCGTGGGTTGGG-ACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGT 777
Db 132903 ACTGTTGGCGTGGGTTGGG-ACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGT 132962
Qy 778 GAAGTGGCGAGGCGGCTTACTTGAATTTGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
Db 132963 GAAATGGGCTGGCGCGGCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCAGG 133022
Qy 838 TGCAATGACCTTAATCGTGGCTCTACTCAATCGCGTGCAGCATTTTGTCCAGCGCGC 897
Db 133023 CGCTATCGATCTGCATACTCTCGCCAGACGCAATCGCGGGGTGGCTGTTCACACGGGC 133082
Qy 898 AGTTTGTGTAATCTACCAATCCCAAGTATTGTGTTTCTGGCGCGCTATTTCGCCGA 957
```

Db 133083 GATATTGTCAATCTAAACCAATCCAAAAGTATGTCTTTCTTCCGCCCTGTTCCGCA 133142  
Qy 958 ATTCAATCATGCCGGAACAGCCGCAACTGATGAGTATATCGTCTCGGCTCACCACATAT 1017  
Db 133143 ATTCAATCATGCCGGAACAGCCGCAACTGATGAGTATATCGTCTCGGCTCACCACAT 133202  
Qy 1018 TGTGTCGATATTATTTGATGATCGGTTACGCCACCTTGTCTCAACGATGCTCTATG 1077  
Db 133203 TGTGTCGATATTATTTGATGATCGGTTACGCCACCTTGTCTCAACGATGCTCTATG 133262  
Qy 1078 GATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTGGCTCGTTTATGCT 1137  
Db 133263 GATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTGGCTCGTTTATGCT 133322  
Qy 1138 GGTGGAGCGCTGTTAGCATCGGCGAGGCATGCGTGAA 1176  
Db 133323 GGTAGCGCGCTCTCTGGCTCGGCAAGACACGCGTGATA 133361

RESULT 12  
AE008884/c  
LOCUS AE008884 21692 bp DNA linear BCT 23-APR-2003  
DEFINITION Salmonella typhimurium LT2, section 188 of 220 of the complete genome.  
ACCESSION AE008884 AE006468  
VERSION AE008884.1 GI:16422503  
KEYWORDS  
SOURCE Salmonella typhimurium LT2  
ORGANISM Salmonella typhimurium LT2  
REFERENCE 1 (bases 1 to 21692)  
AUTHORS McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.  
TITLE Complete genome sequence of Salmonella enterica serovar Typhimurium LT2  
JOURNAL Nature 413 (6858), 852-856 (2001)  
MEDLINE 21534948  
PUBMED 11677609  
REFERENCE 2 (bases 1 to 21692)  
AUTHORS  
CONSTRM The Salmonella typhimurium Genome Sequencing Project  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA  
COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>  
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>  
The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; [http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)  
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

FEATURES  
source  
Location/Qualifiers  
1. .21692  
/organism="Salmonella typhimurium LT2"  
/mol\_type="genomic DNA"  
/strain="LT2; SGSC 1412; ATCC 700720"  
/db\_xref="ATCC:700720"  
/db\_xref="taxon:99287"  
/note="LT2"  
110. .2843  
/gene="cyaA"  
/note="synonym: STM3939"  
110. .118  
/gene="cyaA"  
/note="putative -35\_signal for cyaA;  
RegulonDB:STMLTH004631"  
136. .151  
/gene="cyaA"  
/note="putative binding site for CRP, RegulonDB: STMS1H000108"  
/bound\_moiety="CRP"  
137. .145  
/gene="cyaA"  
/note="putative -10\_signal for cyaA;  
RegulonDB:STMLTH004631"  
284. .289  
/gene="cyaA"  
/note="putative RBS for cyaA; RegulonDB:STMS1H003879"  
297. .2843  
/gene="cyaA"  
/EC number="4.6.1.1"  
/note="adenylate cyclase. (SW:CYAA\_SALTY)"  
/codon\_start=1  
/transl\_table=11  
/product="adenylate cyclase"  
/protein\_id="AAL22784.1"  
/db\_xref="GI:16422504"  
/translation="MYLYIETLKQRLDAINQLRVDRALAMGPAFQQVYSLPTLLHY  
HPLMPGYLDGNVPSGICFYTPDETQRHYNELEYRGMTPODPPKGLPITGVYTMG  
STSSVQSCSDLDIWHQSWLDGEEROLLQKCSLLESWAASLGVEVSFFLIDENR  
FRHNSGSLGGDCGTQHILLDEFYRTAVRLAGKRLMSVPCDEEHYDDYVMTL  
YAQGLTPNEWLDLGLSSLSAEYFGASLWLYKSIDSPYKAVLKTLLLEAYSWEYP  
NPRLLAKOIKQRLHDGEIVSFGLDPCNMLEIVTEYLTAEIDPTRLDLVRCFYLKVC  
EKLRSERACVGRREVLSQVSEWGDARLTMLDRANWKIDQVREAHNELLDAMQ  
SYRNLIREFARNNLSVASPDQIGVLRKLYAAFEALPGKVTLVNPOISPDSEPNLT  
FIHVPPGRANRSGWLYNRPANMDSIIHQPLEYRNLYNKLVAWAFNGLLTSRTHLF  
IKNGIVDLPLQEMVADVSHHFPRLRLPAPTKALYSPCEIRHLAIIVNLEYDPTAAF  
RNKVVFDFRKLDFVSGEONCLIGSIDLLYRNWNEVRTLHFNGEQAMIEALKTIL  
GKMHQDAAPDSVEVFCYSQHLRGLIRTRVQQLVSECLRLSSTRQTRFKALRVS  
GQTWGLFFERLNVQKLENAIEFYGAISHNKLHGLSVQVETNQKLPSPVVDGFASEG  
IIQFFETGDEKGFNIYILDESNAEYVHCEGSKELVRDVSRYSSSHDRFTYGS  
SPINFNLPOFYIVKTDGRAQVIPFRTQPIINTVPPANQDHDAPLLOQYFS"  
2943. .3304  
/gene="STM3940"  
2943. .2948  
/gene="STM3940"  
/note="putative RBS for STM3940; RegulonDB:STMS1H003880"  
2951. .3304  
/gene="STM3940"  
/codon\_start=1  
/transl\_table=11  
/product="putative inner membrane protein"  
/protein\_id="AAL22785.1"  
/db\_xref="GI:16422505"  
/translation="MADAFILLGIVVMVSLGFIILINKLFCFISAGCLSLCASMASF  
QLWDASYWGRWGKVCPLGLDVIISCDNYHFLYDLGWELYGIAFLFTALMLTCAAILI  
NMIMALERYCAGWRR"  
3486. .4032  
/gene="STM3941"  
3486. .3491  
/gene="STM3941"  
/note="putative RBS for STM3941; RegulonDB:STMS1H003881"  
3490. .4032  
/gene="STM3941"

```

/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22786.1"
/db_xref="GI:16422506"
/translaltion="MDVFIKGYMSINNPILICVYNKKRQIKLALFYSVVAIALMWSLF
LNYISMLKMWCVFLILLMIAGASAYWYSAFSGKPLQTLNQGVTLHTRLPPIVYWHIEI
DYVGERVSDNTPVLAIFVKDVELYQCRITNEKMRNNFSLNKKHESNRVMNISLNDLD
YDSDELQDIFKNAVARNLEQ"
4040. .4406
/gene="STM3942"
RBS
4040. .4045
/gene="STM3942"
/note="putative RBS for STM3942; RegulonDB:STMS1H003882"
4044. .4406
/gene="STM3942"
/codon_start=1
/transl_table=11
/product="putative cytoplasmic protein"
/protein_id="AAL22787.1"
/db_xref="GI:16422507"
/translaltion="MFFVITHHIDVLPLATLPHTHETCPYCNNDQVWLIIRQKRTRTCG
LSQARKRDXKFGIAICNHCSEIKEKRWSPALRQLFTQKPLFNLTFWQRYGFVWGLS
WVPALFLATWLTYPTFRGW"
complement(4477. .4810)
/gene="cyay"
/note="CYAY protein. (SW:CYAY_SALTY)"
complement(4477. .4797)
/gene="cyay"
/note="CYAY protein. (SW:CYAY_SALTY)"
/codon_start=1
/transl_table=11
/product="putative frataxin family transport protein"
/protein_id="AAL22788.1"
/db_xref="GI:16422508"
/translaltion="MNDSEHRLADALWLTIEERLDSWDGSDIDCEINGVLTLSFE
NGSKIIINRQEPHQVWLATKQGGYHFDLKGDWVCDRSGETFDLLEQAATQOAGEK
VSFR"
complement(4805. .4810)
/gene="cyay"
/note="putative RBS for cyay; RegulonDB:STMS1H003883"
complement(4867. .5262)
/gene="STM3944"
complement(4867. .5256)
/gene="STM3944"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22789.1"
/db_xref="GI:16422509"
/translaltion="MPKMTPLNRFALHKLKALRGVIGISCIYVIESIMHGTING
VEAAGVLTISAVLGLLSALLDIKAKPRAPLILLHFLLCGGCVALLMVAMLALNHPFVN
ASVILITGALFIVLYLPLSLFYLSRQR"
complement(5244. .5766)
/gene="STM3945"
/note="pseudogene; 2 frameshifts"
/pseudo
complement(5257. .5262)
/gene="STM3944"
/note="putative RBS for STM3944; RegulonDB:STMS1H003884"
5946. .6155
/gene="yifL"
/note="synonym: STM3946"
5946. .5951
/gene="yifL"
/note="putative RBS for yifL; RegulonDB:STMS1H003886"
5952. .6155
/gene="yifL"
/codon_start=1
/transl_table=11
/product="putative outer membrane lipoprotein"
/protein_id="AAL22790.1"
/db_xref="GI:16422510"

```

```

/translation="MKNVFKTLAVLLTLFSLTGCGLKGPLYFPFPADKNAPPPTKKVDS
QTQSTMPDKNDRATGDGPSQVNY"
6180. .7016
gene
Query Match 62.7%; Score 753; DB 1; Length 21692;
Best Local Similarity 78.5%; Pred. No. 3.2e-223;
Matches 926; Conservative 0; Mismatches 250; Indels 3; Gaps 2;
QY 1 AGAAATAATGTGGAGATCCGACCGCCCATCGAATGTCAGTATATAGCGTTTACGCCAC 60
Db 20235 AGGAACAGGGTGGCTATCGCGCCCCCAATTGAGTGGCCAGGATATAACGCTTACGCCAC 20176
QY 61 GGACCGGGCTGAACCTCTCTGCTGCCAGAAATCCGCCAGATCATCAACATAATCAATAAG 120
Db 20175 GGGCCCGGCTCAATTTCTCTGCTGCCAGAACGCCGCTAAGTCTCTCAACATAATCGTTAAA 20115
QY 121 CGATTACATGCCCGAGATCGGCTAACAGGGACCCGGAACGTCCTCCCTGCCCGCGA 180
Db 20115 TGATCGACATGACCCCGATCGGGTCCGATAACATCCGCCCGGAGCGCCCTTGACCACGG 20056
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCGGCATAT 240
Db 20055 TGGTCGATGATGAAGATATCAAAACCGAGATGAATAAGGTCTATACGCCAGTTCAGCGTAT 19996
QY 241 TTTACGTAGCTCTCAATACGCCCGCGGAGATGACTACACCCGGTTCATGGTGTGCTGCG 300
Db 19995 TTTACATAACTCTCAATGCGGCCAGGACAAATGACGATCGTCCGATCGTTGGAATCATTT 19936
QY 301 CGAAAAACGGACAAAGCGCACCGGAATGTTCATCCACACAGTAACTCTGCTTCATCACGC 360
Db 19935 CGAAAAACGGACAAAGCGCACCGGAATGTTCATCCACCGGATAACTCCGCCCTCTTCCCGC 19876
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 420
Db 19875 TGACGCCAGAAATCGTACGCGGCCCCATAGCAAAAGCGCAACACGATTTTCTCTTGT 19816
QY 421 TCCAGTCTTTTGTCTGCTGAAACATCGGGTAAATCTGCTCTTAAACACGTAATAATCGT 480
Db 19815 TCCAGTCAITTTGTCTGCTGAAACATCGGGTAAATCTGCTCTTAAACAGTAAATCGT 19756
QY 481 TTTTNTTAGCGTGCCTGACACAAACGCTGCGACAGTAGCGTATTGTGCAACAAAAATAGAC 540
Db 19755 TTTTNTGCGTGCCTGACACAAACAGCGGAGTAGCGTATTGTGCAACAAAAACAGAC 19696
QY 541 ACACCGGGAGTTCATCATGACCTTAGAATGCTGTTGCTTCCCTACCTGCTGACATCATCAT 600
Db 19695 AACACGGAGTTCGCGATGACCTTTGAATGGTGGTTCCGCTATCTACTGACCTCAACTCT 19636
QY 601 TTTAACGCTGTGCCAGGCTCTGGTGCATCAACACTATGACCACTCGCTCAACACCGG 660
Db 19635 GCTGAGTCTTTCTCCGGGTTACGGGCCCATCAATACCATGACGACGCTCTATCAACCATGG 19576
QY 661 TTATCCGGCCG--GTGGCGTCTATTGCTGGGCTTCAGACCGGACTGCGGAGTTCATATTGT 718
Db 19575 ATATCGTGGCGAGTGGCTTCTATCGCCGACTCCAGACCGGGCTGGGATACATATCGT 19516
QY 719 GCTGGTTGGCGTGGGTTGGG--ACGCTATTTTCCCGTTCAGTGATTCGCTTTGAAGTGT 777
Db 19515 ACTGGTGGCGTGGACTGGGTACGCTCTTTTTCGCGTTCGCTCCTCGCTTTTGAATCT 19456
QY 778 GAAGTGGCGAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
Db 19455 GAAATGGCTGGCGCGGCTTATCTTATCTGGCTGGGTATCCAGCAATGGCGCGCGCAGG 19396
QY 838 TGCAATTGACCTTAAATCGCTGGCTCTACTCAATCGGTCGACATTTGTTTCCAGCGCGC 897
Db 19395 CGCTATCGATCTGCATACCTCGCCAGACGCAATCGGGGGTCCGGTGTTCAAACGGGC 19336
QY 898 AGTTTGTGAATCTCACCAATCCCAAGATATTGTGTTTCTGGCGGCGCTATTTCGCA 957
Db 19335 GATATTGTCAATCTAACCAATCCCAAGATATTGTGTTTCTTGTCCGCGCTTTTCGCA 19276
QY 958 ATTCAATCATGCGGCAACAGCGCAACTGATGATGATATATATGCTGCTCGCGCTCACCCTAT 1017

```



|||||  
Db 19275 ATTCATCATGCCGACGCACTGGCGAGTACCTCATTCTGGCGTACCACGAT 19216  
|||||  
QY 1018 TGTGGTCGATATTATTGTGATGATCGGTTACGCCACCTTGCCTCAACGGATTGCTCTATG 1077  
|||||  
Db 19215 TGTGGTGGATATGTTGTGATGATACCGGTTACGCCACACTGGCGAGCGCATTCGCCGCTG 19156  
|||||  
QY 1078 GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCT 1137  
|||||  
Db 19155 GATTAAGGACCAAGCAGATGAAGCGCTGAATAAGCGTTTGGTTCGTTGTTATGCT 19096  
|||||  
QY 1138 GGTGGAGCGCTGTTAGCATCGGCGAGGCGATCGGTGAAA 1176  
|||||  
Db 19095 GGTAGGCGGCTCCTGGCGTCGGCAAGACACGCGTGATA 19057  
|||||

RESULT 13  
STYSTMD1/c  
LOCUS STYSTMD1 96086 bp DNA linear BCT 11-FEB-2000  
DEFINITION Salmonella typhimurium fragment STMD1.  
ACCESSION AF233324  
VERSION AF233324.1 GI:6960215  
KEYWORDS  
SOURCE  
ORGANISM Salmonella typhimurium LT2  
Salmonella typhimurium LT2  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
REFERENCE 1 (bases 1 to 96086)  
Washington University Genome Sequencing Center.  
The Salmonella typhimurium Genome Sequencing Project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 96086)  
Waterston,R.  
Direct Submission  
AUTHORS  
TITLE Submitted (09-FEB-2000) Department of Genetics, Washington  
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
JOURNAL  
COMMENT  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: sclifton@watson.wustl.edu or  
jspieth@watson.wustl.edu

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality  
>= 30); an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from manually evaluated  
computer analysis, using similarity information and the programs  
GLIMMER (Salzberg, S., Delcher, A., Kasif, A. and White, O. (1998)  
NAR 26,544-548), and GeneMark (Lukashin, A.V. and Borodovsky, M.  
(1998), NAR 26,1107-1115.

FEATURES  
source

Location/Qualifiers  
1..96086  
/organism="Salmonella typhimurium LT2"  
/mol\_type="genomic DNA"  
/strain="SGSC1412"  
/db\_xref="taxon:99287"  
/clone="STMD1"  
<1..980  
/product="23S ribosomal RNA"  
/note="rrlC"  
1072..1191  
/product="5S ribosomal RNA"  
/note="rrfC"  
1244..1323

rRNA

rRNA

tRNA

/product="tRNA-Asp"  
1329..1401  
/product="tRNA-Trp"  
complement (1506..2354)  
/gene="STMD1.99"  
complement (1506..2354)  
/gene="STMD1.99"  
/note="83% identity with E. coli hypothetical protein  
(YIFA) (SP:P22788) and 95% identity with amino acids 1-80  
of E. coli possible regulatory protein (PSSR) (SP:P27826)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/protein\_id="AAF33488.1"  
/db\_xref="GI:6960298"  
/translation="MESTVDTLLKTFLEVSRTTRHFGRAAEALYLTQSAVSFRIQLE  
NQLGVNLFTRHNNIRLTAGEKLLPYAETIMNTWQAARKEVAHTSRHNFSGASAS  
LWECMLNWLGRLYQLQEQSLQFEARIAQROSLVKQLHERQLDLLITTEAPKMD  
SSQLGHFTTALYCSPARKSELYLRLWGPDPFQQHETGLIAADEVPVLTSSAEL  
ARQQLSALNGCSWLPVNWANEKGLHTVADSATLSRPLVAILWQNSDKYSLICDLIKT  
DVLDEQ"  
2461..2799  
/gene="yife"  
2461..2799  
/gene="yife"  
/note="STMD1.96"  
/codon\_start=1  
/transl\_table=11  
/product="95% identity with E. coli hypothetical protein  
(YIFE) (SP:P27827)"  
/protein\_id="AAF33485.1"  
/db\_xref="GI:6960295"  
/translation="MAESFTTNRYFDNKHYPGRGSRHGDFTTKEAQLERHGHAFND  
LDLGRPEVTEEEKLFVAVCRGERPVTDAERVWSKYMTRIKPKRFHTLSGGKPQVE  
GADYTEADD"  
complement (2824..4344)  
/gene="yifB"  
complement (2824..4344)  
/gene="yifB"  
/note="STMD1.95"  
/codon\_start=1  
/transl\_table=11  
/product="85% identity to E. coli hypothetical protein  
(YIFB) (SP:P22787); contains similarity to Pfam family  
PF01078 (Magnesium chelatase, subunit ChlI), score=351.8,  
E=7.6e-102, N=1"  
/protein\_id="AAF33484.1"  
/db\_xref="GI:6960294"  
/translation="MSLAIVHTRAALGVNAPPITIEVHISNGLPLTMVGLPETTVKE  
ARDVRSAIINSGYEFPKAKITINLAPADLPKEGGVDLPPIAVALAASEQLTASNLE  
AYELVGEALTALGALVPGAISSATEAIRAGNIIIVATENAAEVLISKEGCFIADHL  
QTVCAFLEKHALERPLAQDVASPTADLRDVIQGGKRGLEITAAGHNLILIGP  
PTGKTMLASRLSGILPPLSNEEALESAAILSLVNADTVQKRQWQRPFRSPHSAIT  
AMVGGGAIPAPGEISLAHNGILFLDELPEFERRTLDALREPIESGGIHLSTRAKITY  
PARFOLIAAMNPSPTGHYQGNHNRCTPEQTLRYLNRLSGPFLDRFDLSLEIPLPPGI  
LSQHASKGESATVKRVIAAHERQYRQKKLNARLEGREIQKYCVLHDDARWLEDT  
LVHLGLSIRAWORLLKVARTIADIELADQISRQHLQEAVSRAIDRLTIHLQKLLA"  
4696..4794  
/gene="ilvL"  
4696..4794  
/gene="ilvL"  
/note="STMD1.92"  
/codon\_start=1  
/transl\_table=11  
/product="100% identity to E. coli ilvGEDA operon leader  
peptide (ilvL) (SP:P03060)"  
/protein\_id="AAF33486.1"  
/db\_xref="GI:6960296"  
/translation="MTALLRVISLVISVVVVIIPPCGAALGRGKA"  
4934..6580  
/gene="ilvG\_1"  
4934..6580  
/gene="ilvG\_1"

gene

CDS

gene

CDS

/note="STMD1.91"  
/codon\_start=1  
/transl\_table=11  
/product="92% identity with E. coli acetolactate synthase  
II (ILVG) (SP:P00892) ; contains similarity to Pfam  
family PF00205 (Thiamine pyrophosphate enzymes),  
score=952.6, E=1.5e-295, N=1"  
/protein\_id="AAF33483.1"  
/db\_xref="GI:6960293"  
/translation="MNGAQMVVHVALRAQGVKTVFGYGGAIMPVVDALYDGGVEHLLC  
RHEQGAAMAALGYARSTGKTVCIATSGPGATNLITGLADALLDSVPVVAITQVVSAP  
FIGTDAFQEVGLSLACTKHSFLVQSLLELPRIVAEAFEVANAGRPGPVLVDIPKD  
IQLASCELEPFTTVANEATFPQADVEQARQMLEQAKKPMLYVGGVGMAQAVPALRK  
FIATVMPVCTLTKGLGAVEADYPYVLGMLGHGTAAANFAVQECOLLIAVGARFDDR  
VTGKLTFAPNASVIHMDIDPAEMNKLQAHVALQGLNSLLPALQQPLKIDAWRQSC  
AEIRAEHAWRYDHPGETIYAPLLKQLSERKPADSVVTTDVGHQHWMSAQHMYTRPE  
NFIITSSGLTGWGFLPAAVGAQVARNPDVICISGDSPFMNVQELGTVKRKQLPLKI  
VLLDNQRLGMVRQWQQLFFQERYSETLTNDPDLMLASAFGIPGGHITRKDQVEAAL  
DTMLASEGPYLLHVSIDELENVWPLVPPGASNSEMLEKLS"  
6580.6840  
/gene="ilvM"  
6580.6840  
/gene="ilvM"  
/note="STMD1.90"  
/codon\_start=1  
/transl\_table=11  
/product="93% identity with E. coli acetolactate synthase  
(ILVM) (SP:P13048)"  
/protein\_id="AAF33482.1"  
/db\_xref="GI:6960292"  
/translation="MQHVNVSARFNPETLERVLRVVRHGRFQVCSMNMEATDAQNI  
NIELTVASPRSDLLFSQLSKLVDVAHVAICQSAATSQIRA"  
6588.7787  
/gene="ilvE"  
6588.7787  
/gene="ilvE"  
/note="STMD1.89"  
/codon\_start=1  
/transl\_table=11  
/product="S. typhimurium branched-chain-amino-acid  
transaminase (ILVE) (SP:P15168)"  
/protein\_id="AAF33481.1"  
/db\_xref="GI:6960291"  
/translation="MTTKADYIWFNGEMVRWEDAKVHVMHALHYGTSVFEGIRCYD  
SHKGPVFRHEHMQRLDSAKIYRFPVVSQSIDELMEACRDVIRKNNLTSAYIRPLVF  
VGDVGVMVNPYPGYTTDVIIAAPWGAYLGAELDQIDAMVSWNRAPNPIPTAAK  
AGNYLSLLVGSSEARRHGYQEGIALDVNGYISEGAGENLFVKGDLFTPPFTSSAL  
PGITRDATIKLAKELGIEVREQLSRESLYLADEVFMSGTAETIPVRSVDGIQVGE  
RCGPVTKRIQQAFFGLFTGETEDKXGWLDPVNS"  
7948.9798  
/gene="ilvD"  
7948.9798  
/gene="ilvD"  
/note="STMD1.88"  
/codon\_start=1  
/transl\_table=11  
/product="S. typhimurium dihydroxy-acid dehydratase (ILVD)  
(SP:P40810) ; contains similarity to Pfam family PF00920  
(Dehydratase family), score=1023.9, E=3.5e-304, N=1"  
/protein\_id="AAF33480.1"  
/db\_xref="GI:6960290"  
/translation="MPKYRSATTHGNMAGARALWRATGMTDSDFGKPIIAVNVNSFT  
QFVPGHVLRLDLGLVAEQIEASGGVAKENFTIAVDGDIAMGHGMLYSLPSRELIAD  
SVEYMNACADAMVCISNCKITPGMLMASRLNIPVIFVSGGPMAGTKLSLSDKII  
KLDLVDMITQADPKVSDQSNQVERSACTCGSCSGMFTANMNCLEALGLSQPN  
GSLLATADRKQLFLNAGKRIVELTKRYEQNDESLPRNIAKAAFENAMTLDIAMG  
GSTNTVHLAAQAEIDFTMSDIDKLSRKVPQLCKVAPSTQKYHMEDEVHAGGVLG  
ILGELDRAGLLNRNVNVLGLTLPLQLEQYDITVQDEAVKKNFRAGPAGIRTTQAFS  
QDCRWSDLDRAAGCIRSLLEYAYSXKDGGLAVLGNFAENGCIKVTAGVDDSLKFTG  
PAKVYESQDAVEAILGKVVGEVDVVIRYEGPKGPGMQEMLYPTSLKSMGLGKAC  
ALITDGRFSGTSLSGHSVPEAASGTIALIEDGTIAIDIPNRSIQLQSLSEAEIA  
ARAEQAEARGDKAWTPKNRQVSVFALRAYASLATSADKXGVRDKSKLGG"

Query Match 62.7%; Score 753; DB 1; Length 96086;  
Best Local Similarity 78.5%; Pred. No. 4.4e-223;  
Matches 926; Conservative 0; Mismatches 250; Indels 3; Gaps 2;  
QY 1 AGAAATAATGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 62264 AGGAACAGGGTGGCTATCGCGCCCCCATGTAGTGCAGGATATAACGCTTACGCCAC 62205  
QY 61 GGACCGGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120  
Db 62204 GGCCCCGGCTCAATTTCTCTGCTGCCAGAACCGCCGTAAGTCTCTCAACATAATCGTTAAAA 62145  
QY 121 CGATTAACATGCCCGAGATCGCGATCGGCTAACAGGCGACCGGAACGTCCTCCCTGCCCGCGA 180  
Db 62144 TGATCGACATGACCCCGATCGGGTCCGATAACATCCGCCCGGAGCGCCCTTGACACCGG 62085  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAAGAGGTCATAGGCCAGTTCGCCATAT 240  
Db 62084 TGGTCGATGATGAAGATATCAAAACCCGAGATGAAAAGGTCATACGCCAGTTCAGCGTAT 62025  
QY 241 TTTACGTAGCTCTCAATACGCCCCCGGGCAGATGACTACCAACCGGTCATGGTGTGTGCG 300  
Db 62024 TTTACATACTCTCAATGCGGCCAGGACAAATGACCATCGTCCGATCGTTGGAATCATTT 61965  
QY 301 CGAAAACGGACAAAGCGCACCGGGAATGTCTCCACACACAGTAAACTCTGCTTCATCACGC 360  
Db 61964 CGAAAACGGACAAAGCGCACCGGGAATGTCTCCACACCGGGAATAACTCCGCTCTTCCCGC 61905  
QY 361 TGACGCCAGAAATCAGTCAGCGGGTCCCATGGTAAAGACAGCAAAACGCGTTTCTCTTGT 420  
Db 61904 TGACGCCAGAAATCGGTAGCGGGCCCATAGCAAAAGCGGCAAAACGCAATTTCTCTTGT 61845  
QY 421 TCCAGTCTTTTGTGCTGTAACACATCGGGTAATCTGCCTCTTAAACACACGTAATAATCGT 480  
Db 61844 TCCAGTCAATTTGCTGTAACACATCGGGTTTTCGGCCTCGTTTACCAGGTAAATCGT 61785  
QY 481 TTTTGTAGCGTCCCTGACACAAACGCTGCGACAGTAGCGTATTGTGGCACAAAAATAGAC 540  
Db 61784 TTTTGTGCTGCTGCTGACACAAACAGCGGCGAGTAGCGTATTGTGGCACAAAAACAGAC 61725  
QY 541 ACACCGGGAGTTCATCATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCAT 600  
Db 61724 AACAGGGAGTTTCGATGACCTTTGAATGGTGGTTGCCTATCTACTGACCTCAACTCT 61665  
QY 601 TTTAACGCTGTGCCAGGCTCTGGTGAATCAACATCAACTATGACCACTCGCTCAACACCG 660  
Db 61664 GCTGAGTCTTTCTCCGGTTTCCGGGCGCATCAATACCATGACGAGCTCTATCAACCATGG 61605  
QY 661 TTATCCGGCGC--GTGCGTCTATTGCTGGGTTTCCAGACCGGACTGGCGATTATATATTGT 718  
Db 61604 ATATCGTGGCGCAGTGGCTTTATCCCGGACTCCAGACCGGGTGGGGATACATATCGT 61545  
QY 719 GCTGTTGGCGTGGGTTGGG-ACGCTATTTTCCGCTCAGTGATTGCTTTGAAGTGT 777  
Db 61544 ACTGTTGGCGTGGGACTGGGTACGCTCTTTTCGGCTCGCTCTCTCGCTTTTGAAATCT 61485  
QY 778 GAAGTGGCGAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGAGTGGCGCGCGCTGG 837  
Db 61484 GAAATGGGCTGGCGCGGCTTATCTTATCTGGCTGGGTATCCAGCAATGGCGCGCGCGCAGG 61425  
QY 838 TGCAATTGACCTTAATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGC 897  
Db 61424 CGCTATCGATCTGCATACCTCGCCAGACGCAATCGCGGGTGGCTGTTTCAACCGGC 61365  
QY 898 AGTTTTTGTGAATCTCAACAAATCCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCA 957  
Db 61364 GATATTTGTCAATCTAACCAATCCCAAAAGTATTGTCTTTCTTCCGCCCTGTTTCCGCA 61305  
QY 958 ATTTCATATCGCGCAACAGCGCAACTGATGCAATATATCGTGTCTGGCGTCAACACTAT 1017  
Db 61304 ATTTCATATCGCGCAACAGCGCAACTGCGCGCACTACCTCATTTCTGGCGTCAACAGAT 61245  
QY 1018 TGTGTCGATATTATTGTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATG 1077

Db 61244 TGTGGTATATGTTGTGATGACCGGTTACGCCACACTGGCGCAGCGCATTCGCCGCTG 61185  
QY 1078 GATTAAAGGACCAAGCAGATGAGCGGCTGAATAAGATTTTCGGCTCGTTGTTTATGCT 1137  
Db 61184 GATTAAAGGACCAAGCAGATGAGCGGCTGAATAAGCGTTGTTGTTGTTTATGCT 61125  
QY 1138 GGTGGAGCGCTGTAGCATCGCGAGGCGATCGCTGAAA 1176  
Db 61124 GGTAGCGCGCTCCTCGCGTCGGCGCAAGACACCGCGTGATA 61086  
RESULT 14  
AC020874  
LOCUS AC020874 265383 bp DNA linear HTG 16-FEB-2000  
DEFINITION Mus musculus clone RP23-333M14, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC020874  
VERSION AC020874.2 GI:6980207  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 265383)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 265383)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 16, 2000 this sequence version replaced gi:6686434.  
\* NOTE: This record contains 244 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1 869: contig of 869 bp in length  
gap of unknown length  
870 1181: contig of 312 bp in length  
gap of unknown length  
1182 2202: contig of 1021 bp in length  
gap of unknown length  
2203 2960: contig of 758 bp in length  
gap of unknown length  
2961 3646: contig of 686 bp in length  
gap of unknown length  
3647 4520: contig of 874 bp in length  
gap of unknown length  
4521 4813: contig of 293 bp in length  
gap of unknown length  
4814 5434: contig of 621 bp in length  
gap of unknown length  
5435 6000: contig of 566 bp in length  
gap of unknown length  
6001 6992: contig of 992 bp in length  
gap of unknown length  
6993 7785: contig of 793 bp in length  
gap of unknown length  
7786 8367: contig of 582 bp in length  
gap of unknown length  
8368 9114: contig of 747 bp in length  
gap of unknown length  
9115 9729: contig of 615 bp in length  
gap of unknown length  
9730 10313: contig of 584 bp in length  
gap of unknown length  
10314 10712: contig of 399 bp in length  
gap of unknown length  
10713 11404: contig of 692 bp in length  
gap of unknown length  
11405 12261: contig of 857 bp in length  
gap of unknown length  
12262 13173: contig of 912 bp in length  
gap of unknown length  
13174 13934: contig of 761 bp in length  
gap of unknown length  
13935 14703: contig of 769 bp in length  
gap of unknown length  
14704 15380: contig of 677 bp in length  
gap of unknown length  
15381 15654: contig of 274 bp in length  
gap of unknown length  
15655 16714: contig of 1060 bp in length  
gap of unknown length  
16715 17249: contig of 535 bp in length  
gap of unknown length  
17250 17863: contig of 614 bp in length  
gap of unknown length  
17864 18508: contig of 645 bp in length  
gap of unknown length  
18509 19193: contig of 685 bp in length  
gap of unknown length  
19194 19737: contig of 544 bp in length  
gap of unknown length  
19738 20600: contig of 863 bp in length  
gap of unknown length  
20601 21239: contig of 639 bp in length  
gap of unknown length  
21240 21903: contig of 664 bp in length  
gap of unknown length  
21904 22705: contig of 802 bp in length  
gap of unknown length  
22706 22948: contig of 243 bp in length  
gap of unknown length  
22949 23682: contig of 734 bp in length  
gap of unknown length  
23683 24364: contig of 682 bp in length  
gap of unknown length  
24365 25288: contig of 924 bp in length  
gap of unknown length  
25289 25954: contig of 666 bp in length  
gap of unknown length  
25955 26094: contig of 140 bp in length  
gap of unknown length  
26095 27015: contig of 921 bp in length  
gap of unknown length  
27016 27740: contig of 725 bp in length  
gap of unknown length  
27741 28060: contig of 320 bp in length  
gap of unknown length  
28061 28343: contig of 283 bp in length  
gap of unknown length  
28344 29016: contig of 673 bp in length  
gap of unknown length  
29017 29371: contig of 355 bp in length  
gap of unknown length  
29372 29586: contig of 215 bp in length  
gap of unknown length  
29587 29691: contig of 105 bp in length  
gap of unknown length  
29692 30501: contig of 810 bp in length  
gap of unknown length  
30502 31132: contig of 631 bp in length  
gap of unknown length  
31133 31249: contig of 117 bp in length  
gap of unknown length  
31250 32467: contig of 1218 bp in length  
gap of unknown length

* 32468	32878:	contig of 411 bp in length
* 32879	34069:	gap of unknown length
* 34070	34694:	contig of 1191 bp in length
* 34695	35933:	gap of unknown length
* 35934	36920:	contig of 625 bp in length
* 36921	37173:	gap of unknown length
* 37174	37282:	contig of 1239 bp in length
* 37283	38196:	gap of unknown length
* 38197	38452:	contig of 987 bp in length
* 38453	38678:	gap of unknown length
* 38679	38835:	contig of 109 bp in length
* 38836	39561:	gap of unknown length
* 39562	40452:	contig of 256 bp in length
* 40453	41223:	gap of unknown length
* 41224	41864:	contig of 726 bp in length
* 41865	42690:	gap of unknown length
* 42691	42988:	contig of 891 bp in length
* 42989	43612:	gap of unknown length
* 43613	44265:	contig of 771 bp in length
* 44266	45263:	gap of unknown length
* 45264	46592:	contig of 641 bp in length
* 46593	47054:	gap of unknown length
* 47055	47193:	contig of 826 bp in length
* 47194	48027:	gap of unknown length
* 48028	48670:	contig of 139 bp in length
* 48671	49511:	gap of unknown length
* 49512	49895:	contig of 834 bp in length
* 49896	50124:	gap of unknown length
* 50125	50776:	contig of 643 bp in length
* 50777	50905:	gap of unknown length
* 50906	51260:	contig of 841 bp in length
* 51261	52029:	gap of unknown length
* 52030	52294:	contig of 129 bp in length
* 52295	53140:	gap of unknown length
Query Match		
Best Local Similarity		
Matches 748; Conservative		
58.2%;	Score 699;	DB 2; Length 265383;
96.3%;	Pred. No. 4.2e-206;	
0;	Mismatches 25;	Indels 4; Gaps 3;

QY	200	CAAACCCCAATGGAACAGGTCTATAGGCCAGTTCGGCATATTTTACGTAGCTCTCAATAC	259
Db	55129	CATAATCCAATGGAACAGGTCTATAGGCCAGTT-CGATATTTTACGTAGCTCTCAATAC	55187
QY	260	GCCCCGGGCAGATGACTACACCCGGTCTATGGTGTCTGCGGAGAAACGGACAAAGCGCA	319
Db	55188	GCCCCGGGCAGATGACTACACCCGGTCTATGGTGTCTGCGGAGAAACGGACAAAGCGCA	55247
QY	320	CGGAATGTCTATCCACACAGTAAACTCTGTCTTATCAGCGTGTACGCGAGAAATCAGTCA	379
Db	55248	CGGAATGTCTATCCACACAGTAAACTCTGTCTTATCAGCGTGTACGCGAGAAATCAGTCA	55307
QY	380	GCGTCCCATGTATAAAGCAGCAACCGGTTTTCTCTGTCTTCCAGTCTTTTGTGCT	439
Db	55308	GCGTCCCATGTATAAAGCAGCAACCGGTTTTCTCTGTCTTCCAGTCTTTTGTGCT	55367
QY	440	GAAACATCGGGTAATCTGCCTCTTAAACACAGTAAATCGTTTTTTAGCGTGCCTGAC	499
Db	55368	GAAACATCGGGTAATCTGCCTCTTAAACACAGTAAATCGTTTTTTAGCGTGCCTGAC	55427
QY	500	ACAACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGACACACCCGGAGTTTCATG	559
Db	55428	ACAACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGACACACCCGGAGTTTCATG	55487
QY	560	ACCTTAGAATGGTGGTTTGCTACCTGTGACATCGATCATTTTAAACGCTGTGCCAGGC	619
Db	55488	ACCTTAGAATGGTGGTTTGCTACCTGTGACATCGATCATTTTAAACGCTGTGCCAGGC	55547
QY	620	TCTGGTGCATCAACACATATGACCACTCGCTCAACACAGGTTAT--CCGGCCCGTGGCG	677
Db	55548	TCTGGTGCATCAACACATATGACCACTCGCTCAACACAGGTTATCGCGCGCGTGGCG	55607
QY	678	TCTATTGCTGGCTTCAGACGGGACTGGGATTCATATTGTGTTGGTGGGTT-736	736
Db	55608	TCTATTGCTGGCTTCAGACGGGACTGGGATTCATATTGTGTTGGTGGGTTG55667	55667
QY	737	GGGACGCTATTTCCCGCTCAGTGATGGGTTTGAAGTGTGAAGTGGCGAGCGCGCT	796
Db	55668	GGGACGCTATTTCCCGCTCAGTGATGGGTTTGAAGTGTGAAGTGGCGAGCGCGCT	55727
QY	797	TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCG	856
Db	55728	TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCG	55787
QY	857	CTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGAGTTTTTGTGAATCTCACC	916
Db	55788	CTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGAGTTTTTGTGAATCTCACC	55847
QY	917	AATCCCAAAAGTATGTTTCTGGCGGCTATTTCCGCAATTCATCATGCCGCAA	973
Db	55848	AATCCCAAAAGTATGTTTCTGGCGGCTAGTCTCGTACCGCCAGGTGCTCGAA	55904
RESULT 15			
AE015396/c			
LOCUS			
DEFINITION			
Shigella flexneri 2a str. 301 section 359 of 412 of the complete genome.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Shigella flexneri 2a str. 301			
Shigella flexneri 2a str. 301			
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.			
REFERENCE			
AUTHORS			
Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H., Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L., Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.			
Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli			

JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
PUBMED 12384590  
REFERENCE 2 (bases 1 to 10253)  
AUTHORS Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y., Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R. China

FEATURES  
source Location/Qualifiers  
1..10253  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
complement(243..2167)  
/gene="yigE"  
/locus\_tag="SF3893"  
/pseudo  
complement(243..2167)  
/gene="yigE"  
/locus\_tag="SF3893"  
/note="Similar to Escherichia coli K12 orf, conserved hypothetical protein gi: 1790248 (162 aa). BLAST with identity of 99% in 641 aa. This CDS has been disrupted by insertion. The sequence has been checked and is believed to be correct."  
/pseudo  
/codon\_start=1  
/transl\_table=11  
472..1818  
/locus\_tag="SF3892"  
472..1818  
/locus\_tag="SF3892"  
/note="Residues 1 to 448 of 448 are 98 pct identical to residues 1 to 448 of a 448 aa protein from Shigella flexneri ref: NP\_085177.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45328.1"  
/db\_xref="GI:24054380"  
/translation="MPPDFMHIGQALDLVSRVDSLRLNPLTSLGDYLDPELISRCLAE SGTVTRKRLPLEMVMWCIVGVLEKEPLHQIVNRDIDMLPGRNPFVAPSAVIOAR QRLGSEAVRRVFTKTAQLWNNATPHPHWCGLTLAIDGVFWRTPDTPENDAAFPRQTH AGNPALYPQVKVCOMELTSHLTAFAAGFTMKNSENELAEOLIEQTGDNTLTLMDDGY YSLGLNGWSLAGEHRHWMIPLRGAQYEEELKLGKGDHLVKLTSPQARKKWPGLGN EVTARLLTVTRKGVCHLLTSMTDAMRFPGEEMADLYSNRWEIELGYREIKQTMOLSR LTLRSKAPELVEQELWGLLAYNLVRYQMIKVAEHLKGYWPNQLSFSSESGMVRMLM TLQASPGRIPELMSDLASMGQLVKLPTRRGRAPPRVVKERPWKYPTAPKKSQVA"  
2522..3487  
/gene="corA"  
/locus\_tag="SF3894"  
2522..3487  
/gene="corA"  
/locus\_tag="SF3894"  
/note="Residues 6 to 321 of 321 are 100 pct identical to residues 1 to 316 of a 316 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_290445.1"  
/codon\_start=1  
/transl\_table=11  
/product="Mg2+ transport, system I"  
/protein\_id="AAN45329.1"  
/db\_xref="GI:24054381"  
/translation="MGVPVMLSATOLENNRLTRLEVESQPLVNAVWIDLVEPDDDER LRVQSELGQSLATRPELEDIEASARFFEDDDGLHIHSFFFEADAEDHAGNSTVAFTIR DGRFLTFRERELPAFLYMRARSQSMVDGNAYVELLLDLFTETKIEQLADEIENIYSDL EQLSRVINEGHQGDYEALSTLAELEDIGKVKRLCLMDTORALNFLVRKARLPGGQL EQAREILRDIESLLPHNESLFFQVNFMLQAMGMFINIEQNRIIKIFSVVSVWFLPPTL

gene VASSYGMNFEFMPPELKWSFGYPGAIIFMILAGLAPLYFKRKNWL"  
complement(3528..3950)  
/gene="yigP"  
/locus\_tag="SF3895"  
CDS complement(3528..3950)  
/gene="yigP"  
/locus\_tag="SF3895"  
/note="Residues 15 to 140 of 140 are 99 pct identical to residues 1 to 126 of a 126 aa protein from Escherichia coli K12 ref: NP\_418261.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45330.1"  
/db\_xref="GI:24054382"  
/translation="MHLIHQNNNSYGVFMSKEYMNDLSLSEKWKYRFNFYDQHGFPGF WGATPEYKAAPKALKVRQLTIQMNFIAPFCSWIYLFVLGLWKKAILVLLGILSLFV GALIGVNIIGIAVAAYAVNTNKFYEKVKGLNTWSL"  
complement(3922..4362)  
/gene="yigG"  
/locus\_tag="SF3896"  
/gene="yigG"  
CDS complement(3922..4362)  
/gene="yigG"  
/locus\_tag="SF3896"  
/note="Residues 9 to 146 of 146 are 95 pct identical to residues 1 to 138 of a 138 aa protein from Escherichia coli K12 ref: NP\_418262.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45331.1"  
/db\_xref="GI:24054383"  
/translation="MYESGHIAMLTGDSHKDVKFMLRMFIPTSNKGISRHRHYIFLFI LNFIFAFLLIFFNDGEAGFLVIVSTIALHYLVINNQCRLRDSGFIYIKIYVFGTLAV YIISITIMIAEDFACSGNSMIFLICYPSTFSLMLAPTSSKQ"  
complement(4397..5287)  
/gene="rard"  
/locus\_tag="SF3897"  
complement(4397..5287)  
/gene="rard"  
CDS /locus\_tag="SF3897"  
/note="Residues 1 to 296 of 296 are 99 pct identical to residues 1 to 296 of a 296 aa protein RARD\_ECOLI sp: P27844"  
/codon\_start=1  
/transl\_table=11  
/product="Rard protein"  
/protein\_id="AAN45332.1"  
/db\_xref="GI:24054384"  
/translation="MDAKQTRQGVLLAALAAFYINGIAPAYFKLIYVVPADAILTHRVI WSFFFMVMSICRQNSYLKTLIQPKIFMLAVSAVLIGNWLLFIWAVNNHMLLEA SLGYFINPLVNIVLGMIFLGERFRMOWLAIVLCGVLVQLWTFGSLPIIALGLAPS FAFYGLVRKKAIAVEAQTMLIETWLLPVAIYLEAIADSSSTSHMCQNPMNLNLLIA AGIVTTVPLLCFTAATRLRLSTLGFQYIGPTLMFLAVTFYGEKPGADKMWITFTFI WVLAIFVMDAIYQRTSK"  
complement(5339..5824)  
/gene="yigI"  
/locus\_tag="SF3898"  
complement(5339..5824)  
/gene="yigI"  
CDS /locus\_tag="SF3898"  
/note="Residues 1 to 161 of 161 are 100 pct identical to residues 1 to 161 of a 161 aa protein from Escherichia coli O157:H7 ref: NP\_312777.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45333.1"  
/db\_xref="GI:24054385"  
/translation="MKCKDDMSAVLTABQALKLVGMFVYHMPFNALGMELEREYEKE FAQLAFKNQPMVGNWQASILHGQVIASALDVAAGLVCGSTLTRHETISEDLRQRL SRMGTDILRDVYLRPGRGERFTATSSLLRAGNKVAVARVELHNEQLYIASATATYV G"



gene 5971. .6840  
/gene="p1dA"  
/locus\_tag="SF3899"  
CDS 5971. .6840  
/gene="p1dA"  
/locus\_tag="SF3899"  
/note="Residues 1 to 289 of 289 are 100 pct identical to residues 1 to 289 of a 289 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_290453.1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane phospholipase A"  
/protein\_id="AAN45334.1"  
/db\_xref="GI:24054386"  
/translation="MRTLQGWLLPVFMLPMAYVAGEATVKEVHDAPAVRGSILANMLQ EHDNFTLYDYDTNLYITQTSDLNKEAIAASYDAENARKDEVKFLSLAPFLMRGIL GPNSVLGASYTQKSWQLSNSESSPRETNYPQLFLGFATDYRFAGWTILRDVEMGY NHDNNGRSDFTSRSWNRLYRLMAENGNWLVEKWPYVVGNTDNDPDIITKYNGYYQLK IGYHLGDAVLSAKQYNNWNTGYGGAELGLSYPIITKHVRLYTVQVSGYGESLIDYNFNQ TRVGVGVMLNDLF"  
gene 6988. .8802  
/gene="recQ"  
/locus\_tag="SF3900"  
CDS 6988. .8802  
/gene="recQ"  
/locus\_tag="SF3900"  
/note="Residues 1 to 604 of 604 are 99 pct identical to residues 8 to 611 of a 611 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_290454.1"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent DNA helicase"  
/protein\_id="AAN45335.1"  
/db\_xref="GI:24054387"  
/translation="MLNDESQAKQVLQETFGYQQFRPGQEEIIDTVLSGRDCLVVMPT GGGKSLCYQIPALLNGLTVVSPVLSLMDKQDQDQLQANGVAACLNSTQTREQQLEV

Query Match 57.5%; Score 690; DB 1; Length 10253;  
Best Local Similarity 98.2%; Pred. No. 1.4e-203;  
Matches 719; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY	472	TAAATCGTTT	TTTACGCTGCGTACACACACGCTGCGACAGTAGCGTATTGTGGCACA	531
Db	10253	TAAATCGTTT	TTTAAAGCGTGCGTACACACACGCTGCGACAGTAGCGTATTGTGGCACA	10194
QY	532	AAATAGACACACCGGAGTT	CATCATGACCTTAGAATGGTGTTCCTACCTGCTGAC	591
Db	10193	AAATAGACACACCGGAGTT	CATCATGACCTTAGAATGGTGTTCCTACCTGCTGAC	10134
QY	592	ATCGATCA	TTTAAACGCTGTCGCCAGGCTCTGTGCAATCAACACTATGACACCTCGCT	651
Db	10133	ATCGATCA	TTTAAAGCCTGTGCGCAGGCTCTGTGCAATCAACACTATGACACCTCGCT	10074
QY	652	CAACCA	CGGTTAT--CGGCCGGTGGCGTCTATTGCTGGGCTCAGACCGGACTGGCGAT	709
Db	10073	CAACCA	CGGTTATCGCGCGCGTGGCGTCTATTGCTGGGCTCAGACCGGACTGGCGAT	10014
QY	710	TCATATTG	CTGTTGGCGTGGGTT-GGGACGCTATTTCCCGCTCAGTGA	768
Db	10013	TCATATTG	CTGTTGGCGTGGGTTGGGACCGCTATTTCCCGCTCAGTGA	9954
QY	769	TGAAGTGT	GAAGTGGCAGCGCGGCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCG	828
Db	9953	TGAAGTGT	AAAGTGGCAGCGCGGCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCG	9894
QY	829	CGCGCTG	GTGCAATTGACCTTAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTT	888
Db	9893	CGCGCTG	GGCAATTGATCTTAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTT	9834
QY	889	CCAGCGG	CAGTTTTTTGTGAATCTCACC	948
Db	9833	CCAGCGG	CAGTTTTTTGTGAATCTCACC	9774

QY	949	ATTTCCGCAATTCAT	CGCCGCAACAGCCGCACTGATGCAGTATATCGTGTCTGGCGT	1008
Db	9773	ATTTCCGCAATTCAT	CGCCGCAACAGCCGCACTGATGCAGTATATCGTGTCTGGCGT	9714
QY	1009	CACCACTATTG	TGCTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGAT	1068
Db	9713	CACCACTATTG	TGCTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGAT	9654
QY	1069	TGCTCTATGGATTAA	AGGACCAAGCAGATGAAGGCGCTGAATAAGATTTTTCGGCTCGTT	1128
Db	9653	TGCTCTATGGATTAA	AGGACCAAGCAGATGAAGGCGCTGAATAAGATTTTTCGGCTCGTT	9594
QY	1129	GTTTATGCTG	TGGAGCGCTGTTAGCATCGCGAGGCGATGCGTGAAAAATAATGTCCGA	1188
Db	9593	GTTTATGCTG	TGGAGCGCTGTTAGCATCGCGAGGCGATGCGTGAAAAATAATGTCCGA	9534
QY	1189	TGCGGCGTAAAC	1200	
Db	9533	TGCGGCGTAAAC	9522	

Search completed: March 28, 2004, 16:44:55  
Job time : 5039.17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 12:14:43 ; Search time 597.686 Seconds  
(without alignments)  
8529.292 Million cell updates/sec

Title: US-09-847-392-1  
Perfect score: 1200  
Sequence: 1 agaaataatgtggagatgc.....atgtcgatgcggcgtaaac 1200

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1200	100.0	1200	3	Aaz94405 E. coli r
2	1200	100.0	1231	3	Aaa48442 E. coli L
3	571.2	47.6	622	5	Aas93395 DNA encod
4	523.6	43.6	2216	5	Aas89415 DNA encod
5	446	37.2	1023	7	Aca32708 Prokaryot
6	417	34.8	417	4	Aah84646 E. coli g
7	417	34.8	417	7	ACA18626 Prokaryot
8	382.2	31.9	621	7	ACA51728 Prokaryot
9	375	31.2	618	7	ACA36103 Prokaryot
10	357.4	29.8	618	7	ACA32157 Prokaryot
11	348.2	29.0	623	7	ACA48809 Prokaryot
12	298.8	24.9	1017	7	ACA19693 Prokaryot
13	297.2	24.8	1017	7	ACA51727 Prokaryot
14	291.4	24.3	2054	5	Aas88021 DNA encod
15	290.8	24.2	990	7	ACA32154 Prokaryot
16	276.4	23.0	990	7	ACA35868 Prokaryot
17	275	22.9	621	7	ACA53970 Prokaryot
18	262.8	21.9	1026	7	ACA48808 Prokaryot
19	259	21.6	259	7	ACA13399 Prokaryot
20	254.2	21.2	4303	5	Aas85519 DNA encod
21	248.6	20.7	1884	5	Aas89412 DNA encod
22	243	20.2	1032	5	Aas85510 DNA encod
23	243	20.2	1746	5	Aas74968 DNA encod

C	24	243	20.2	2410	5	AAS88022	Aas88022 DNA encod
C	25	207.4	17.3	263	4	AAH84479	Aah84479 E. coli g
C	26	206	17.2	2410	5	AAS88022	Aas88022 DNA encod
C	27	205.2	17.1	1032	5	AAS85510	Aas85510 DNA encod
C	28	181.8	15.2	1014	7	ACA53968	Aca53968 Prokaryot
C	29	175	14.6	274	5	AAS89410	Aas89410 DNA encod
C	30	167.8	14.0	516	5	AAS93398	Aas93398 DNA encod
C	31	149.8	12.5	518	7	ACA52910	Aca52910 Prokaryot
C	32	144.8	12.1	1002	7	ACA44944	Aca44944 Prokaryot
C	33	141	11.8	1041	5	AAS93397	Aas93397 DNA encod
C	34	133	11.1	110000	7	ACF67367_37	Continuation (38 o
C	35	133	11.1	110000	7	ACF67367_38	Continuation (39 o
C	36	133	11.1	110000	7	ACF65388_10	Continuation (11 o
C	37	132.4	11.0	999	7	ACF70714	Acf70714 Photorhab
C	38	112.4	9.4	630	7	ACA42760	Aca42760 Prokaryot
C	39	104.2	8.7	1746	5	AAS74968	Aas74968 DNA encod
C	40	104	8.7	630	7	ACA45284	Aca45284 Prokaryot
C	41	100	8.3	100	7	ACD71734	AcD71734 E. coli K
C	42	100	8.3	100	7	ACD71736	AcD71736 E. coli K
C	43	100	8.3	100	7	ACD71735	AcD71735 E. coli K
C	44	86.8	7.2	633	7	ACA43829	Aca43829 Prokaryot
C	45	80	6.7	627	7	ACA27048	Aca27048 Prokaryot

ALIGNMENTS

RESULT 1

AAZ94405 ID AAZ94405 standard; DNA; 1200 BP.

XX AC AAZ94405;

XX DT 18-JUL-2000 (first entry)

XX DE E. coli rhtB gene imparting homoserine resistance.

XX KW Homoserine resistance; rhtB gene; L-homoserine; L-alanine; L-isoleucine;

XX KM L-valine; L-threonine; ss.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

FT CDS 557..1171

FT /\*tag= a

FT /note= "a DNA sequence corresponding to the coding region (minus the stop codon) is specifically claimed in Claim 3"

XX EP994190-A2.

PN 19-APR-2000.

XX 20-SEP-1999; 99EP-00118581.

XX 13-OCT-1998; 98RU-00118425.

XX (AJIN ) AJINOMOTO CO INC.

XX Livshits VA, Zakataeva NP, Aleoshin VW, Belareova AV; Tokhmakova IL;

XX WPI; 2000-273530/24.

XX P-PSDB; AAY79298.

XX Novel RhtB protein, useful for generation of L-homoserine resistance in Escherichia bacteria and large-scale production of e.g. L-homoserine and L-alanine.

XX Claim 3; Page 10-11; 14pp; English.

XX This is the DNA sequence the novel rhtB gene of Escherichia coli K-12. The rhtB gene maps at 86 min on the E. coli chromosome. The gene

CC participates in resistance to homoserine. Amplification of the gene  
CC results in an improvement of the amino acid productivity of E. coli. The  
CC invention provides: the RhtB protein (see AAY79298); DNA encoding the  
CC RhtB, especially nucleotides 557-1171 of the present sequence; a  
CC bacterium, especially of the genus Escherichia, in which L-homoserine  
CC resistance is enhanced by amplifying the copy number or increasing the  
CC expression rate of the rhtB DNA, the DNA being carried on a multicopy  
CC vector or on a transposon; and a method for producing an amino acid by  
CC cultivating the bacterium in a culture medium to produce and accumulate  
CC the amino acid in the medium, from which it is recovered. The method is  
CC used for the production of L-homoserine, L-alanine, L-isoleucine, L-  
CC valine or L-threonine (all claimed)  
XX  
SQ Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1200; DB 3; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60  
Db 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60  
QY 61 GGACCGGCTGAACCTCTGCTGCCAGAAATGCCGCCAGATCATCAACATATCAATTAAG 120  
Db 61 GGACCGGCTGAACCTCTGCTGCCAGAAATGCCGCCAGATCATCAACATATCAATTAAG 120  
QY 121 CGATTAAACATGCCCCGAGATCGCGTAAACAGCGGACCGGAACGTCCTGCCCGCGA 180  
Db 121 CGATTAAACATGCCCCGAGATCGCGTAAACAGCGGACCGGAACGTCCTGCCCGCGA 180  
QY 181 TGGTCGATGATTAAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 181 TGGTCGATGATTAAGACATCAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
QY 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTPACCAACCCGTCATGTGCTGTGG 300  
Db 241 TTTACGTAGTCTCAATACGCCCGCGGAGATGACTPACCAACCCGTCATGTGCTGTGG 300  
QY 301 CGAAAACGGACAAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCATCAGC 360  
Db 301 CGAAAACGGACAAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCATCAGC 360  
QY 361 TGACGCCAGAAATCAGTACAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTTTGTT 420  
Db 361 TGACGCCAGAAATCAGTACAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTTTGTT 420  
QY 421 TOCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTCTTAAACACAGTAAATCGT 480  
Db 421 TOCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTCTTAAACACAGTAAATCGT 480  
QY 481 TTTTGTAGCGTGCCTGACACACCGTCCGACAGTAGCGTATTGTGGCACAAATAGAC 540  
Db 481 TTTTGTAGCGTGCCTGACACACCGTCCGACAGTAGCGTATTGTGGCACAAATAGAC 540  
QY 541 ACACCGGAGTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCATCAT 600  
Db 541 ACACCGGAGTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCATCAT 600  
QY 601 TTAAACGCTGTCGCCAGGCTGTTGTCATCAACACTATGACCACTCGCTCAACACCG 660  
Db 601 TTAAACGCTGTCGCCAGGCTGTTGTCATCAACACTATGACCACTCGCTCAACACCG 660  
QY 661 TTATCCGGCCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGGATTCATATTGTC 720  
Db 661 TTATCCGGCCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGGATTCATATTGTC 720  
QY 721 TGGTTGGCGTGGGTTGGACGCTATTTCCCGCTCAGTGATTGCGTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGACGCTATTTCCCGCTCAGTGATTGCGTTGAAGTTGAA 780  
QY 781 GTGGCAGGCGCGGCTTACTTGATTGGCTGGGAATCCACAGTGGCGCGCGCTGGTGC 840  
Db 781 GTGGCAGGCGCGGCTTACTTGATTGGCTGGGAATCCACAGTGGCGCGCGCTGGTGC 840

Db 781 GTGGCAGGCGCGGCTTACTTGATTGGCTGGGAATCCACAGTGGCGCGCGCTGGTGC 840  
QY 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGGTCGACATTTGTTCCAGCGCAGT 900  
Db 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGGTCGACATTTGTTCCAGCGCAGT 900  
QY 901 TTTTGTGAATCTACCAATCCCAAGATATTGTTTCTGGCGGCTATTTCCGCAATT 960  
Db 901 TTTTGTGAATCTACCAATCCCAAGATATTGTTTCTGGCGGCTATTTCCGCAATT 960  
QY 961 CATCATGCCGCAACAGCGCAACTGATGTCAGTATATCGTCTGGCGTCACCACTATTGT 1020  
Db 961 CATCATGCCGCAACAGCGCAACTGATGTCAGTATATCGTCTGGCGTCACCACTATTGT 1020  
QY 1021 GGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTCTCAACGATTTGCTCTATGGAT 1080  
Db 1021 GGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTCTCAACGATTTGCTCTATGGAT 1080  
QY 1081 TAAAGGACCAAGCAGATGAAGCGGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAGCAGATGAAGCGGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
QY 1141 GGGAGCGCTTTAGCATCGCGGAGGCGATCGTGAATAAATAATGTCGATCGCGCTAAAC 1200  
Db 1141 GGGAGCGCTTTAGCATCGCGGAGGCGATCGTGAATAAATAATGTCGATCGCGCTAAAC 1200  
RESULT 2  
AAA48442  
ID AAA48442 standard; DNA; 1231 BP.  
XX AAA48442;  
AC AAA48442;  
DT 08-SEP-2000 (first entry)  
XX  
DE E. coli L-homoserine resistance gene, rhtB.  
XX  
KW L-homoserine resistance; L-homoserine synthesis; rhtB; L-threonine;  
KW L-valine; L-leucine; ds.  
XX Escherichia coli.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 557..1174  
FT /\*tag= a  
FT /product= "RhtB"  
XX  
PN EP1013765-A1.  
XX  
PD 28-JUN-2000.  
XX  
PF 20-DEC-1999; 99EP-00125406.  
XX  
PR 23-DEC-1998; 98RU-00123511.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
PI Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;  
XX  
DR WPI; 2000-414602/36.  
DR P-PSDB; AAY99597.  
XX  
PT Novel Escherichia bacterium having enhanced L-threonine resistance due to  
PT enhanced RhtC protein activity, used to produce L-threonine, L-  
PT homoserine, L-valine and L-leucine.  
XX  
PS Claim 4; Page 12-13; 24pp; English.  
XX  
CC The present sequence is the L-homoserine resistance gene, rhtB, from  
CC Escherichia coli. This sequence may be used to impart L-homoserine  
CC resistance on E. coli bacteria, which would be useful for producing a  
CC high yield of L-homoserine. L-homoserine resistance means that the  
CC bacteria will be able to grow on a minimal medium containing L-homoserine

CC at a concentration at which the corresponding wild-type strain would not  
CC grow. Since the transformed bacteria can grow on the minimal medium, it  
CC can synthesise L-homoserine, which accumulates. The accumulated amino  
CC acids can then be removed from the culture medium. The bacterium of the  
CC present invention may also be used to synthesise L-threonine, L-valine  
CC and L-leucine at increased levels  
XX

SQ Sequence 1231 BP; 291 A; 319 C; 311 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 1200; DB 3; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGTCGACCGCCCATCGAATGTGCCAGTATATAGCTTTACGCCAC 60  
Db 1 AGAATAATGTGGAGTCGACCGCCCATCGAATGTGCCAGTATATAGCTTTACGCCAC 60  
QY 61 GGACCGGCTGAACCTCTGCTGCCAGAAATGCCGCCAGATCATCAACATATTAAG 120  
Db 61 GGACCGGCTGAACCTCTGCTGCCAGAAATGCCGCCAGATCATCAACATATTAAG 120  
QY 121 CGATTAAATGCGCGAGATGCGGATCGGCTAAACAGGCGACCGGAACGTCCTGCCCGGA 180  
Db 121 CGATTAAATGCGCGAGATGCGGATCGGCTAAACAGGCGACCGGAACGTCCTGCCCGGA 180  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGAACAGGTCATAGGCCAGTCCGCATAT 240  
Db 181 TGGTCGATGATTAAGACATCAAAACCCCAATGAACAGGTCATAGGCCAGTCCGCATAT 240  
QY 241 TTTACGTAGCTCTCAATACGCCCCCGGCGAGATGACTACACCGGTCATGCTGTGG 300  
Db 241 TTTACGTAGCTCTCAATACGCCCCCGGCGAGATGACTACACCGGTCATGCTGTGG 300  
QY 301 CGAAACCGGACAAAGCGCACCGGATGTCTATCCACACAGTAACTCTGCTTCATCAGC 360  
Db 301 CGAAACCGGACAAAGCGCACCGGATGTCTATCCACACAGTAACTCTGCTTCATCAGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACGCTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTCTTAAACCCACGTAATCGT 480  
Db 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAATCTGCCTCTTAAACCCACGTAATCGT 480  
QY 481 TTTTGTAGCGTGCCTGACACAAACGCTCGACAGTACGTAATGTGGCACAAAATAGAC 540  
Db 481 TTTTGTAGCGTGCCTGACACAAACGCTCGACAGTACGTAATGTGGCACAAAATAGAC 540  
QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCAT 600  
Db 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCAT 600  
QY 601 TTTAACGCTGTCGCCAGGCTCTGTCGAATCAACACTATGACCACTCGCTCAACCCAGG 660  
Db 601 TTTAACGCTGTCGCCAGGCTCTGTCGAATCAACACTATGACCACTCGCTCAACCCAGG 660  
QY 661 TTATCCGCGCGGTGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATTTGTC 720  
Db 661 TTATCCGCGCGGTGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATTTGTC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTCCCGTCAGTGAATGCGTTGAAGTGTGA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTCCCGTCAGTGAATGCGTTGAAGTGTGA 780  
QY 781 GTGGGAGGCGGCTTACTTGTGCTGGGATCCAGAGTGGCGCGCTGGTGC 840  
Db 781 GTGGGAGGCGGCTTACTTGTGCTGGGATCCAGAGTGGCGCGCTGGTGC 840  
QY 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGGTCGACATTTGTTCCAGCGGCAGT 900  
Db 841 AATTGACCTTAAATCGTGGCTCTACTCAATCGGTCGACATTTGTTCCAGCGGCAGT 900

QY 901 TTTTGTGAATCTCACCATCCCAAGATATTGTGTTCTGGCGGCTATTTCGCAATT 960  
Db 901 TTTTGTGAATCTCACCATCCCAAGATATTGTGTTCTGGCGGCTATTTCGCAATT 960  
QY 961 CATCATGCCGCAACAGCGCAACTGATGTCAGTATATCGTCTCGGCTCACCACATTTGT 1020  
Db 961 CATCATGCCGCAACAGCGCAACTGATGTCAGTATATCGTCTCGGCTCACCACATTTGT 1020  
QY 1021 GGTCGATATTATTGTGATGATCGGTTACGCCACCCCTGCTCAACGATTTGCTATGGAT 1080  
Db 1021 GGTCGATATTATTGTGATGATCGGTTACGCCACCCCTGCTCAACGATTTGCTATGGAT 1080  
QY 1081 TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTATGCTGGT 1140  
QY 1141 GGGAGCGCTGTAGCATCGCGGAGGATCGGTGAAATAATATGTCGGATCGCGCTAAAC 1200  
Db 1141 GGGAGCGCTGTAGCATCGCGGAGGATCGGTGAAATAATATGTCGGATCGCGCTAAAC 1200

RESULT 3

AAS93395  
ID AAS93395 standard; cDNA; 622 BP.

XX

AC AAS93395;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #29199.

XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

XX

PR 23-AUG-2000; 2000US-00649167.

XX

(HYSE-) HYSEQ INC.

XX

Drmanac RT, Liu C, Tang YT;

XX

WPI; 2001-639362/73.

DR

P-PSDB; ABG29208.

XX

New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.

XX

Claim 1; SEQ ID NO 29199; 103pp; English.

PS

The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (ii). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal  
activity of (II) or to treat disease states involving (II). (II) is  
useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 622 BP; 123 A; 152 C; 172 G; 175 T; 0 U; 0 Other;  
Query Match 47.6%; Score 571.2; DB 5; Length 622;  
Best Local Similarity 98.1%; Pred. No. 8.7e-179;  
Matches 610; Conservative 0; Mismatches 8; Indels 4; Gaps 3;  
QY 557 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAA-CGCTGTGCGC 615  
Db 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAAGCCCGTCCCC 60  
QY 616 AGGCTCTGGTCAATCAACACTATGACCACTCGCTCAACACGCGTTAT--CGGCCCGGT 673  
Db 61 AGGCTCTGGTCAATCAACACTATGACCACTCGCTCAACACGCGTTATCGCGGCGCGGT 120  
QY 674 GCGCTCTATTGCTGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTGGCGTGG 733  
Db 121 GCGCTCTATTGCTGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTGGCGTGG 180  
QY 734 GTT-GGGACGCTATTTCCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGCGGCGC 792  
Db 181 GTTGGGACGCTATTTCCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGCGGCGC 240  
QY 793 GGCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCGCGCGTGGTGAATTGACCTTAA 852  
Db 241 GGCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCGCGCGTGGTGAATTGACCTTAA 300  
QY 853 ATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCT 912  
Db 301 ATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCT 360  
QY 913 CACCAATCCCAAAAGTATTGTTCTGGCGGCGCTATTTCCGCAATTCAATCATGCCGCA 972  
Db 361 CACCAATCCCAAAAGTATTGTTCTGGCGGCGCTATTTCCGCAATTCAATCATGCCGCA 420  
QY 973 ACAGCCGCAACTGATCAGTATATCGTGTGCTCGGCGTCACCACTATTGTGTCGATATTAT 1032  
Db 421 ACAGCCGCAACTGATCAGTATATCGTGTGCTCGGCGTCACCACTATTGTGTCGATATTAT 480  
QY 1033 TGTGATATCGGTTACCCACCCCTGCTCAACGGATTGCTATGGATTAAAGGACCAAA 1092  
Db 481 TGTGATATCGGTTACCCACCCCTGCTCAACGGATTGCTATGGATTAAAGGACCAAA 540  
QY 1093 GCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTT 1152  
Db 541 GCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTT 600  
QY 1153 AGCATCGCGGAGGCATCGGTGA 1174  
Db 601 AGCATCGCGGAAGCATCGGTGA 622  
RESULT 4  
AAS89415/c  
ID AAS89415 standard; cDNA; 2216 BP.  
XX  
AC AAS89415;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #25219.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.

XX WO200175067-A2.  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG25228.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 25219; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2216 BP; 518 A; 562 C; 639 G; 497 T; 0 U; 0 Other;  
Query Match 43.6%; Score 523.6; DB 5; Length 2216;  
Best Local Similarity 91.3%; Pred. No. 1.3e-162;  
Matches 623; Conservative 0; Mismatches 49; Indels 10; Gaps 6;  
QY 527 GCACAAAATAGACACACCGGAGTTCATCATGACCTTAGAATGGTGGTTCCTACCTG 586  
Db 2116 GCACAAAATAGACAC-CCCGGAGTTCATCATG-CCTTAGAAGGGGGTGGCCACCCGC 2059  
QY 587 CTGACATCGATCATTTTAACGCTGTCGCCAGGCTCTGGTGAATCAACACTATGACCA 646  
Db 2058 GACATTGATCTTTTATTAGCCCGTTCCCGAGTTTCTGGCCCATCACCACCTATGCC 1999  
QY 647 TCGCTCAACACCGGTTAT--CCGGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTG 704  
Db 1998 TCGCTCAACACCGGTTATCGCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCGGACTG 1939  
QY 705 GCGATTTCATATTGCTGGTGGTGGCGTGGGGTT-GGGACGCTATTTCCCGCTCAGTGATT 763  
Db 1938 GCGATTTCATATTGCTGGTGGTGGCGTGGGGTTGGGGACGCTATTTTCCGCTCAGTGATT 1879  
QY 764 GCGTTTGAAGTGTGAAGTGGGCGCGGCTTACTTACTTGAATTTGGCTGGGAATCCAGCAG 823  
Db 1878 GCGTTTGAAGTGTGAAGTGGGCGCGGCTTACTTACTTGAATTTGGCTGGGAATCCAGCAG 1819  
QY 824 TGGCGCGCGCTGGTGAATTTGACCTTAAATCGCTGGGCTCTACTCAATCGCGTCGACAT 883



Db 1818 TGGCGCGCGCTGGTGAATTAATGCTTAAATCGCTGGCTTACTCAATCGCGTGACAT 1759  
QY 884 TTGTTCCAGCG----CGCAGTTTTTGTGAATCTCACCAATCCCAAAGTATTGTTTCT 939  
Db 1758 TTGTTCTACAGCGGCACATTTTTTTGAATCTCTCACCATTCCAAAGTATTGTTTCT 1699  
QY 940 GCGCGGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGCGAGTATATCGT 999  
Db 1698 GCGCGGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGCGAGTATATCGT 1639  
QY 1000 -GCTCGGCGTCAACCACTATTGTTGTCGATATTATTGTTGATCGGTTACGCCACCTTG 1058  
Db 1638 GCGTCGCGCTCACCACTATTGTTGTCGATATTATTGTTGATCGGTTACGCCACCTTG 1579  
QY 1059 CTCAACGAGTTGCTTATGGAATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 1118  
Db 1578 CTCAACGAGTTGCTTATGGAATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 1519  
QY 1119 TCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGCATGCGTGAAAA 1178  
Db 1518 TCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGCATGCGTGAAAA 1459  
QY 1179 TAATGTCGGATGCGCGGTAAC 1200  
Db 1458 TAATGTCGGATGCGCGGTAAC 1437

RESULT 5

ACA32708/C  
ID ACA32708 standard; DNA; 1023 BP.  
XX ACA32708;  
AC ACA32708;  
XX  
DT 19-JUN-2003 (first entry)  
DE Prokaryotic essential gene #14365.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Escherichia coli.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
PI  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU28838.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 20578; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1023 BP; 219 A; 261 C; 289 G; 254 T; 0 U; 0 Other;

Query Match 37.2%; Score 446; DB 7; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 5.1e-137;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTACGCCAC 60  
Db 446 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTACGCCAC 387  
QY 61 GGACCGGGCTGAACCTCTGCTGCCAGAATGCCCGAGATCATCAACATATCATTAAG 120  
Db 386 GGACCGGGCTGAACCTCTGCTGCCAGAATGCCCGAGATCATCAACATATCATTAAG 327  
QY 121 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGCGGACCGGAACGTCCCTGCCCGCA 180  
Db 326 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGCGGACCGGAACGTCCCTGCCCGCA 267  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTAGGCCAGTTCCGCATAT 240  
Db 266 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCTAGGCCAGTTCCGCATAT 207  
QY 241 TTTACGTAGCTCTCAATACGCCCGGGCAGATGACTACCCCGGTCTATGGTGTGCG 300  
Db 206 TTTACGTAGCTCTCAATACGCCCGGGCAGATGACTACCCCGGTCTATGGTGTGCG 147  
QY 301 CGAAACCGGACAAAGCGCACCGGAATGTCTATCCACACCAAGTAACTCTGCTTCATCACGC 360  
Db 146 CGAAACCGGACAAAGCGCACCGGAATGTCTATCCACACCAAGTAACTCTGCTTCATCACGC 87  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACCGGTTTCTTTGTT 420  
Db 86 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACCGGTTTCTTTGTT 27  
QY 421 TCCGAGTCTTTTTTGTGCTGAACAT 446  
Db 26 TCCGAGTCTTTTTTGTGCTGAACAT 1

RESULT 6  
AAH84646  
ID AAH84646 standard; DNA; 417 BP.  
XX  
AC AAH84646;  
XX

DT 26-SEP-2001 (first entry)  
XX E. coli growth and proliferation related coding sequence SEQ ID NO:274.  
DE  
XX Escherichia coli; growth; proliferation; microbial; antimicrobial;  
KW bacterial infection; microorganism; ds.  
XX  
OS Escherichia coli.  
XX  
FN WO200134810-A2.  
XX  
PD 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US030950.  
PF  
XX 09-NOV-1999; 99US-0164415P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX Forsyth RA, Ohlsen K, Zyskind J;  
PI  
XX WPI; 2001-335933/35.  
DR  
DR P-PSDB; AAG98975.  
XX  
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
PT for screening for homologous genes and for designing expression vectors.  
XX  
XX Claim 9; Page 351; 522pp; English.  
PS  
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli growth  
CC and proliferation related proteins given in AAG99078 and AAG98830 to  
CC AAG98999. (I) can be used as potential targets for the generation of new  
CC antimicrobial agents, and for identification of compounds which interact  
CC with the gene products of (I). In addition the expression of (I) and the  
CC purification of the proteins, the purified proteins can be used to  
CC generate reagents and screen small molecule libraries or other candidate  
CC compound libraries for compounds that can be further developed to yield  
CC novel antimicrobial compounds. In addition, nucleic acid probes  
CC complementary to (I) that are specific for particular species of  
CC microorganisms can be used to identify particular microorganism species  
CC in clinical specimens, therefore, providing a rapid and dependable method  
CC by which to identify the causative agents of a bacterial infection. Also,  
CC antibodies generated against proteins translated from mRNA transcribed  
CC from proliferation-required sequences can also be used to screen for  
CC specific microorganisms that produce such proteins in a species-specific  
CC manner. AAH84371 and AAH84670 represent sequencing primers used in the  
CC isolation of E. coli growth and proliferation related sequence, which are  
CC used in an example from the present invention  
XX  
SQ Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match 34.8%; Score 417; DB 4; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.3e-127;  
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 GTGATTGCGTTGAAGTGTGAAGTGGGAGGCGCGGCTTACTTGTATTGGCTGGGAATC 817  
DB 1 GTGATTGCGTTGAAGTGTGAAGTGGGAGGCGCGGCTTACTTGTATTGGCTGGGAATC 60  
QY 818 CAGCAGTGGCGCGCGCGTGGTGCAATTGACCTTAATCGTGGCTCTACTCAATCGCGT 877  
DB 61 CAGCAGTGGCGCGCGCGTGGTGCAATTGACCTTAATCGTGGCTCTACTCAATCGCGT 120  
QY 878 CGACATTGTTCCAGCGCGCAGTTTGTGAATCTACCAATCCCAAAAGATTGTGTTT 937  
DB 121 CGACATTGTTCCAGCGCGCAGTTTGTGAATCTACCAATCCCAAAAGATTGTGTTT 180  
QY 938 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGAGTATATC 997  
DB 181 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGAGTATATC 240  
QY 998 GTGCTCGGCGTCACCACACTATTGTGGTGCATATTATTGTGATGATCGGTTACGCCACCCCTT 1057

DB 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTT 300  
QY 1058 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 1117  
DB 301 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360  
QY 1118 TTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGTAGCATCGGCCAGGCATGCGTGA 1174  
DB 361 TTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGTAGCATCGGCCAGGCATGCGTGA 417

RESULT 7  
ACA18626  
ID ACA18626 standard; DNA; 417 BP.  
XX  
AC ACA18626;  
DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #283.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Escherichia coli.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU14756.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 6496; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match 34.8%; Score 417; DB 7; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.3e-127;  
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 758 GTGATTGGTTGAAGTGTGAAGTGGGCGGCGGCTTACTTGAATTGGCTGGGAATC 817  
Db 1 GTGATTGGTTGAAGTGTGAAGTGGGCGGCGGCTTACTTGAATTGGCTGGGAATC 60  
QY 818 CAGCAGTGGCGGCGGCTGGTGAATTGACCTTAATCGTGGCTCTACTCAATCGGT 877  
Db 61 CAGCAGTGGCGGCGGCTGGTGAATTGACCTTAATCGTGGCTCTACTCAATCGGT 120  
QY 878 CGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTCACCATCCCAAAGTATTGTTT 937  
Db 121 CGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTCACCATCCCAAAGTATTGTTT 180  
QY 938 CTGGCGGCGCTATTCCGCAATTCATCATGCGGCAACAGCGCAACTGATGCAGTATATC 997  
Db 181 CTGGCGGCGCTATTCCGCAATTCATCATGCGGCAACAGCGCAACTGATGCAGTATATC 240  
QY 998 GTGCTCGGCGTACCACTATTGCGTCAATATTGTGATGATCGGTTACGCCACCTT 1057  
Db 241 GTGCTCGGCGTACCACTATTGCGTCAATATTGTGATGATCGGTTACGCCACCTT 300  
QY 1058 GCTCAACGGATTGCTCTATGGATTAAAGACCAAGCAGATGAAGCGCTGAATAAGATT 1117  
Db 301 GCTCAACGGATTGCTCTATGGATTAAAGACCAAGCAGATGAAGCGCTGAATAAGATT 360  
QY 1118 TTCGCTCGTTGTTTATGCTGGCGGAGCGGTGTAGCATCGCGGAGCATCGGTGA 1174  
Db 361 TTCGCTCGTTGTTTATGCTGGCGGAGCGGTGTAGCATCGCGGAGCATCGGTGA 417

RESULT 8  
ACA51728  
ID ACA51728 standard; DNA; 621 BP.  
XX  
AC ACA51728;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #33385.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS *Salmonella typhi*.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
PI  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU47858.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 39598; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 621 BP; 121 A; 170 C; 171 G; 159 T; 0 U; 0 Other;

Query Match 31.9%; Score 382.2; DB 7; Length 621;  
Best Local Similarity 78.1%; Pred. No. 6.3e-116;  
Matches 485; Conservative 0; Mismatches 133; Indels 3; Gaps 2;  
QY 557 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAATTTAACGCTGCGCA 616  
Db 1 ATGACCTTAGAATGGTGGTTTGCCTATCTACTGACCTCAACTCTGTGAGTCTTCTCCG 60  
QY 617 GGCTCTGGTGAATCAACACTATGACCACTGCTGCTCAACCGGTTATCCGCCG--GTG 674  
Db 61 GGTTCAAGCGCCATCAATACCATGACGACGCTCTATCAACCATGATATCGTGGCGCAGCG 120  
QY 675 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTGCTGGTGGCGTGGG 734  
Db 121 GCTTCTATCGCCGGGCTCCAGACCGGCTGGGGATACATATCGTACTGGTGGCGGTCGGA 180  
QY 735 TTGGG--ACGCTATTTCCCGCTCACTGATTTGCGTTTGAAGTGTGAAGTGGGCGGCGG 793  
Db 181 CTGGGTACGCTCTTTTTCGGCTCGCTCATCGCTTTTGAATTTCTGAATGGGCTGGCGG 240  
QY 794 GCTTACTGTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATGACCTTAA 853  
Db 241 GCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCTATCGATCTGCAT 300



Db 601 GCCTCCGCCGCCATGC 617

RESULT 10

ACA32157  
ID ACA32157 standard; DNA; 618 BP.

XX ACA32157;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #13814.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

OS Enterobacter cloacae.

XX WO200277183-A2.

PN 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

PF 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WP1; 2003-029926/02.

DR P-PSDB; ABU28287.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 20027; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 618 BP; 106 A; 176 C; 183 G; 153 T; 0 U; 0 Other;

Query Match 29.8%; Score 357.4; DB 7; Length 618;  
Best Local Similarity 75.9%; Pred. No. 1.1e-107;  
Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY 557 ATGACCTTAGAATGGTGGTTTGCCCTACCTGCTGACATCGATCATTTTAAACGCTGTCCGCA 616  
DB 1 ATGACCTTGAATGGTGGTTTGCCCTACCTGCTGACATCCATTATCTCCAGCCCTTCTCCG 60

QY 617 GGCTCTGTGTCAATCAACACATATGACACACCTCGCTCAACACCGGTTA--TCCGGCCGGTG 674  
DB 61 GGCTCGGGGGCTATTAAACACCATGACACACCTCCATCAATCACGGCTACCGCGGTGCGCG 120

QY 675 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGGACTCATATTGCTGTTGGCGT-GGG 733  
DB 121 GCATCGATTGCCGGTTTGAGACCGGGCTGGTATTTCATATTGCTGGTGGGATCGGT 180

QY 734 GTTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGAGCGCGG 793  
DB 181 CTGGGGACCCCTGTTCTCCCGCTCCGCTGCTGGCCCTTGAGGTGCTAAATGGCGCGGGCA 240

QY 794 GCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAAA 853  
DB 241 GCGTATCTGATTTGGCTCGGCATCCAGCAGTGGCGCGCGGCTCCATCAACCTGAAT 300

QY 854 TCGCTGGCTCTACTCAATCGGTGCGACATTGTTCCAGCGCGGAGTTTTTTGTGAATCTC 913  
DB 301 ACGTGGCTCTGACGCGAAACCGCGCCATCTGTTAAGCGTGCGGTATTCGTCACACCTG 360

QY 914 ACCAATCCCAAAAGTATGTGTTTCTGGCGCGCGCTATTTCCGCAATTCATCATGCCGCA 973  
DB 361 ACCAACCAGAGAGCATCGTGTTCCTCGCGCGCTGTTCCCGGAGTTTATCGTTCGCGAT 420

QY 974 CAGCGGCACTGATGCAATATATCGTCTCGGCGTCCAGTCCACTATGTTGGTGCATATTATT 1033  
DB 421 CAGCCTCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATC 480

QY 1034 GTGATGATCGGTTACGCCACCCCTTGCTCAACCGGATGCTCTATGATTAAGGACCAAG 1093  
DB 481 GTGATGATGTTTACGCCACCGCTGGCGGAGCAATGCGCGGTGATTAAGGCGCTAAG 540

QY 1094 CAGATGAGCGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA 1153  
DB 541 CAGATGAGCGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTT 600

QY 1154 GCATCGGCGAGGCATGC 1170

DB 601 GCGTCAGCGCGTCACGC 617

RESULT 11

ACA48809

ID ACA48809 standard; DNA; 623 BP.

XX ACA48809;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #30466.

DE Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Salmonella paratyphi.

XX WO200277183-A2.

XX 03-OCT-2002.

XX



PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU44939.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 36679; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 623 BP; 124 A; 171 C; 168 G; 160 T; 0 U; 0 Other;  
Query Match 29.0%; Score 348.2; DB 7; Length 623;  
Best Local Similarity 76.6%; Pred. NO. 1.3e-104;  
Matches 479; Conservative 0; Mismatches 138; Indels 8; Gaps 4;  
QY 557 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATCGATCAATTAACGGTCTGCCA 616  
Db 1 ATGACCTTTGAATGGTGGTTGCTGCTATCTACTGACCTCAACTCTGCTGAGTCTTCTCCG 60  
QY 617 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACGGTTATCCGGCG--GTG 674  
Db 61 GGTTCAGGGGCCATCAATACCATGACGACGTCTATCAACCATGGATATCGTGGCGCAGCG 120  
QY 675 GGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATATTTGCTGGTGGTGGGG 734  
Db 121 GCTTCTATCGCGGACTCCAGACCGGACTGGGGATACATATCGTACTGTTGGCGCTCGGA 180  
QY 735 TTGGG-ACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGCGCGCG 793

Db 181 CTGGGTACGCTCTTTTCGCGCTCGCTCTCGTTTGAATTTCTGAATGGGCTGGCGG 240  
QY 794 GCTTACTTGTATTTGGTGGGAATCCAGAGTGGCGCGCTGGTGCATTTGACCTTAA 853  
Db 241 GCTTATCTTATCTGGTGGGTATCCAGCAATGGCG--CGCAGGCGCTATCGATTCGAT 297  
QY 854 TCGCTGGCTCTACTCAATCGCGTCCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 913  
Db 298 ACTCTCGCCAGACGCAATCGCGGGTGGCTGTTCAACGGGGGATTTTGTCAATCTA 357  
QY 914 ACCAATCCCAAAAGTATTGTTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGCA 973  
Db 358 ACCAATCCCAAAAGTATTGTTGTTTCTTTCGCGCGCTGTTTCGCAATTCATCATGCCGCA 417  
QY 974 CAGCCGCAACTGATGCAGTATATCGTGTCTCGGCGTCCACCACTATTGTGGTTCGATATT 1033  
Db 418 CAACCGCAACTGGCGCAGTACCTCATTTCTCGGCGTCCACCACTATTGTGGTGGATGATT 477  
QY 1034 GTGATGATCGGTTACGCCACCTTGTCTCAACGG--ATTGCTCTATGGAATTAAGGACCAA 1091  
Db 478 GTGATGACCGGTTACGCCACACTGGCGCTAGCGCACTTGGCGCTGATTAAGGACCAA 537  
QY 1092 AGCAGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGT 1151  
Db 538 AGCAGATGAAGCGCGCTGAATAAGCGTTTGGTTGTTTATGCTGGTGGCGCGCTCC 597  
QY 1152 TAGCATCGGCGAGGCATGCGTGAAA 1176  
Db 598 TGGCGTCCGCAAGACACGCGTGATA 622  
RESULT 12  
ACA19693/c  
ID ACA19693 standard; DNA; 1017 BP.  
XX ACA19693;  
AC ACA19693;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #1350.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Salmonella typhimurium.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU15823.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 7563; 1766pp; English.  
XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1017 BP; 208 A; 251 C; 305 G; 253 T; 0 U; 0 Other;

Query Match 24.9%; Score 298.8; DB 7; Length 1017;  
Best Local Similarity 79.4%; Pred. No. 4.6e-88;  
Matches 354; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 AGAAATATGTGGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 446 AGGAACAGGTGGCTATCGCGCCCGCCCATTTGAGTGGCCAGGATATAACGTTACGCCAC 387  
QY 61 GGACCGGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAATATTAAG 120  
Db 386 GGCCCGGGCTCAATTTCTCTGCTGCCAGACCGCGCTAAGTCTCAACATAATCGTTAAA 327  
QY 121 CGATTACATGCTCCGAGATCGGATCGGCTAACAGCGGACCGGACGTCCTCCCGCGA 180  
Db 326 TGATCGACATGACCCCGATGCGGGTCCGATAACATCCGCCCGGAGCGCCCTTGACCCAG 267  
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCAATAGGCCAGTCCGCATAT 240  
Db 266 TGGTCGATGATGAAGATATCAAAACCGAGATGAAGGATCATACGCCAGTTCAGCGTAT 207  
QY 241 TTTACGTAGCTCTCAATACGCCCGCGGAGATGACTACCAACCGGTCAATGCTGTGCG 300  
Db 206 TTTACATACTCTCAATGCGCGCCAGGACAAATGACGATCGTCCGATCGTTGGAATCATTT 147  
QY 301 CGAAACCGGACAAAGCGCACCGGAATGTCATCCACACAGTAACTCTGCTTCATCACGC 360  
Db 146 CGAAACCGGACAAAGCGCACCGGAATGTTACCCACCGCATAACTCCGCTCTCCCGC 87  
QY 361 TGACGCCGAAATCAGTCAGCGGTCCTCATGGTAAAGCAGCAACCGGTTTCTTTGTT 420  
Db 86 TGACGCCGAAATCGGTGAGCGGCCCATAGCAAAAGCGGAAACGCAATTTCTCTTTGTT 27  
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446  
Db 26 TCCAGTCTATTTGCTGCTGTAACAT 1

RESULT 13  
ACAS1727/c

ID ACAS1727 standard; DNA; 1017 BP.  
XX  
AC ACAS1727;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #33384.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
OS *Salmonella typhi*.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU47857.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 39597; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1017 BP; 208 A; 249 C; 305 G; 255 T; 0 U; 0 Other;

Query Match 24.8%; Score 297.2; DB 7; Length 1017;  
Best Local Similarity 79.1%; Pred. No. 1.6e-87;  
Matches 353; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 446 AGGAACAGGGTGGCTATCGCGCCCCCATTTGAGTGGCCAGGATATAACGCTTACGCCAC 387  
QY 61 GGACCGGGCTGAACCTCCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120  
Db 386 GACCCCGGCTCAATTTCTGCTGCCAGAACCGCGCTAAGTCTCAACATAATCGTTAAA 327  
QY 121 CGATTAACTGCCCGAGATCGCGGATCGGCTAACAGGCGACCGGACGTCCTCGCCGCGA 180  
Db 326 TGATCGACATGACCCCGATCGCGGTCCGATAACATCCGCCCGGAGCGCCCTTGACACCG 267  
QY 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTATAGCCAGTTCGCGCATAT 240  
Db 266 TGGTCGATGATGAAGATATCAAAACCGAGATGAAAAGGTATACGCCAGTTCAGCGTAT 207  
QY 241 TTTACGTAGTCTCAATACGCCCCCGGCGAGATGACTACCACCGGTTCATGTTGTGCG 300  
Db 206 TTTACATAACTCTCAATGCGGCCAGGACAAATGACGATCGTCCGATCGTTGGAGTCATTC 147  
QY 301 CGAAACCGACAAAGCGCACCGGAATGTCTATCCACACAGTAACTCTGCTTCTATCACGC 360  
Db 146 CGAAACCGACAAAGCGCACCGGAATGTCTATCCACCGCGATAACTCCGCTCTTCCCGC 87  
QY 361 TGACGCCAGAAATCAGTCAGCGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420  
Db 86 TGACGCCAAATAATCGGTACGCGGCCCATAGCAAAAGCGGCAACCGCATTTTCTTGT 27  
QY 421 TCCAGTCTTTTGTGCTGCTGAAACAT 446  
Db 26 TCCAGTCATTTTGTGCTGCTGAAACAT 1

RESULT 14  
AAS88021/c  
ID AAS88021 standard; cDNA; 2054 BP.

AC AAS88021;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #23825.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX

OS Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG23834.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID NO 23825; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 2054 BP; 477 A; 527 C; 563 G; 487 T; 0 U; 0 Other;  
Query Match 24.3%; Score 291.4; DB 5; Length 2054;  
Best Local Similarity 99.7%; Pred. No. 2e-85;  
Matches 292; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 460 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 401  
QY 61 GGACCGGGCTGAACCTCCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120  
Db 400 GGACCGGGTTGAACCTCCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 341  
QY 121 CGATTAAACATGCCCGAGATCGCGATCGGCTAAACAGGCGACCGGAACGTCCTGCCCGGA 180  
Db 340 CGATTAAACATGCCCGAGATCGCGATCGGCTAAACAGGCGACCGGAACGTCCTGCCCGGA 281  
QY 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 280 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 221  
QY 241 TTTACGTAGTCTCTCAATACGCCCGCGGCGAGATGACTACCACCGGTCTATGGTG 293  
Db 220 TTTACGTAGTCTCTCAATACGCCCGCGGCGAGATGACTACCACCGGTCTATGGTG 168

RESULT 15  
ACA32154/c  
ID ACA32154 standard; DNA; 990 BP.

XX  
XX ACA32154;  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX  
DE Prokaryotic essential gene #13811.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Enterobacter cloacae.  
XX  
XX WO200277183-A2.  
PN  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX  
PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU28284.

XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX  
PS Claim 14; SEQ ID NO 20024; 1766pp; English.

XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 990 BP; 187 A; 266 C; 309 G; 228 T; 0 U; 0 Other;

Query Match 24.2%; Score 290.8; DB 7; Length 990;  
Best Local Similarity 78.3%; Pred. No. 2.1e-85;  
Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 446 AAAACAGCGTCGAATGGCGCGCCCATGGAGTGGCAAGAAATATACCGTTGCGCCAG 387  
QY 61 GGACCGGGCTGAACCTCTGCTGCCAGAAATGCCCGCAGATCATCAACATAATCAAG 120  
Db 386 GGGCCGGGCTGAACGTCCTGCTGCCAGAAACCGCGCAAAATCGTCGACGTAAATCGTGAAG 327  
QY 121 CGATTAAACATGCCCCGAGATGCGGATCGGCTAACAGGGCGACCGGAACGTCCTGCCCGGA 180  
Db 326 CGATCGAGTGGCCCGCGGTGGTGTCCGGCAGCATACGTCCGGACAGCCCTGTCCACGA 267  
QY 181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTATAGGCCAGTTCGGCATAT 240  
Db 266 TGGTCGATGATCAGGACATCGAACCCCATATGACCAGGTATAGGCCAGTTCGGCGTAT 207

QY 241 TTTACGTAGTCTCAATACGCCCCCGGCGAGATGACTACCAACCCGGTCATGGTGTGCG 300  
Db 206 TTAACGTAGCTTTCAATACGCCCCCGGCGAGACCAATCAACCCGGTCATTTTTCATTT 147  
QY 301 CGAAAACGGACAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCAACGC 360  
Db 146 CGGAAACGAAACGAAACCGCACCGGATGCCCCCGACGCGCGTAAACTCATCTTCCACGC 87  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAACAGCAGCAAAACCGGTTTCTCTGTT 420  
Db 86 TGCCGCCAGAAATCGGTACGCGGCCCATAGAGAAACAGCAGCAAAACCGGTTTCTCTGTT 27  
QY 421 TCCAGTCTTTTGTGCTGAAACAT 446  
Db 26 TCCAGTCTTTTGTGCTGAAACAT 1

Search completed: March 28, 2004, 14:37:59  
Job time : 603.686 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:10:49 ; Search time 120.331 Seconds  
(without alignments)  
5534.263 Million cell updates/sec

Title: US-09-847-392-1  
Perfect score: 1200  
Sequence: 1 agaaataatggagatgc.....atgtcgatcgcgctaaac 1200

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1200	100.0	1200	4	US-09-396-357-1
2	417	34.8	417	4	US-09-711-164-274
3	391.8	32.6	645	4	US-09-489-039A-895
C 4	297.2	24.8	1041	4	US-09-489-039A-851
C 5	207.4	17.3	263	4	US-09-711-164-107
C 6	144.8	12.1	1038	4	US-09-543-681A-238
7	112.4	9.4	834	4	US-09-252-991A-6447
C 8	102.4	8.5	930	4	US-09-252-991A-6103
9	76.6	6.4	663	4	US-09-328-352-777
10	59	4.9	645	4	US-09-489-039A-2622
11	46	3.8	1830121	4	US-09-557-884-1
12	46	3.8	1830121	4	US-09-643-990A-1
13	37.2	3.1	633	4	US-09-328-352-2682
14	36	3.0	312	4	US-09-543-681A-3779
15	35.4	2.9	1028	4	US-09-655-908-17
C 16	34.4	2.9	612	4	US-09-489-039A-1357
C 17	34.4	2.9	1431	4	US-09-489-039A-1316
C 18	33.8	2.8	3534	4	US-09-134-001C-2269
C 19	33.2	2.8	789	4	US-09-252-991A-2459
20	33.2	2.8	1296	4	US-09-252-991A-2216
21	33.2	2.8	1941	4	US-09-252-991A-2378
C 22	33.2	2.8	2820	4	US-09-252-991A-2538
C 23	33.2	2.8	10993	4	US-08-961-527-15
C 24	33.2	2.8	4403765	3	US-09-103-840A-2
C 25	33.2	2.8	4411529	3	US-09-103-840A-1
26	32.8	2.7	1442	4	US-09-634-238-82
C 27	32.8	2.7	16075	3	US-09-096-942-1

C	28	32.8	2.7	16075	3	US-09-096-867-1	Sequence 1, Appli
	29	32.6	2.7	968	4	US-09-023-655-1112	Sequence 1112, Ap
	30	32.4	2.7	474	4	US-09-621-976-18033	Sequence 18033, A
C	31	32.4	2.7	161652	4	US-09-497-855A-40	Sequence 40, Appli
	32	32.4	2.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	33	32.4	2.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	34	32.2	2.7	681	4	US-09-252-991A-15039	Sequence 15039, A
C	35	32	2.7	1001	4	US-09-641-638-278	Sequence 278, App
C	36	32	2.7	1597	4	US-09-634-238-195	Sequence 195, App
C	37	31.8	2.6	738	4	US-09-252-991A-6015	Sequence 6015, Ap
	38	31.8	2.6	1128	4	US-09-252-991A-6376	Sequence 6376, Ap
	39	31.8	2.6	2136	4	US-09-252-991A-6294	Sequence 6294, Ap
C	40	31.8	2.6	3948	4	US-09-252-991A-6175	Sequence 6175, Ap
	41	31.6	2.6	25356	4	US-09-976-594-750	Sequence 750, App
	42	31.4	2.6	455	4	US-09-621-976-9043	Sequence 9043, Ap
C	43	31.4	2.6	13987	2	US-08-804-227C-13	Sequence 13, Appli
C	44	31.4	2.6	44377	2	US-08-804-227C-7	Sequence 7, Appli
C	45	31.4	2.6	44377	2	US-08-804-198-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-396-357-1  
; Sequence 1, Application US/09396357  
; Patent No. 6303348  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD  
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/396,357  
; CURRENT FILING DATE: 1999-09-15  
; EARLIER APPLICATION NUMBER: RU98118425  
; EARLIER FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (557)..(1171)  
US-09-396-357-1

Query Match	100.0%;	Score 1200;	DB 4;	Length 1200;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1200;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTTACGCCAC	60	
Db	1	AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTTACGCCAC	60	
QY	61	GGACCGGCTGAACCTCTGCTGCCAGAATGCCCGAGATCAACATAATCAATAAG	120	
Db	61	GGACCGGCTGAACCTCTGCTGCCAGAATGCCCGAGATCAACATAATCAATAAG	120	
QY	121	CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTGCCCGCGA	180	
Db	121	CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTGCCCGCGA	180	
QY	181	TGGTCGATGATTAGACATCAACCCCAATGGACAGGTCAATAGGCCAGTTCCGCATAT	240	
Db	181	TGGTCGATGATTAGACATCAACCCCAATGGACAGGTCAATAGGCCAGTTCCGCATAT	240	
QY	241	TTTACGTAGCTCTCAATACGCCCGGCGCAGATGACTACCAACCCGGTCATGGTGTGTGG	300	



Db 241 TTTACGTAGCTCTCAATACGCCCGCCGCGAGATGACTACCAACCGGTCTATGCTGTGCG 300  
QY 301 CGAAACCGACAAAGCGACCGGAATGTCTATCCACCAAGTAACTCTGCTTCATCAGC 360  
Db 301 CGAAACCGACAAAGCGACCGGAATGTCTATCCACCAAGTAACTCTGCTTCATCAGC 360  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420  
Db 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420  
QY 421 TCCAGTCTTTTGTGTGTAACATCGGGTAATCTGCCTCTTAAACCAACGTAATAATCGT 480  
Db 421 TCCAGTCTTTTGTGTGTAACATCGGGTAATCTGCCTCTTAAACCAACGTAATAATCGT 480  
QY 481 TTTTGTAGCGTCCCTGACACAAACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540  
Db 481 TTTTGTAGCGTCCCTGACACAAACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540  
QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACTGCTGACATCAT 600  
Db 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACTGCTGACATCAT 600  
QY 601 TTTAACGCTCTGCCAGGCTCTGTGCAATCAACATATGACCACTCGCTCAACCAACGG 660  
Db 601 TTTAACGCTCTGCCAGGCTCTGTGCAATCAACATATGACCACTCGCTCAACCAACGG 660  
QY 661 TTATCCGCGGCTGCGCTATTGCTGGCTTTCAGACCGGACTGCGGATTCATATTGTC 720  
Db 661 TTATCCGCGGCTGCGCTATTGCTGGCTTTCAGACCGGACTGCGGATTCATATTGTC 720  
QY 721 TGGTTGGCGTGGGTTTGGACCGCTATTTCGCGCTCAGTGATTGGCTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTTGGACCGCTATTTCGCGCTCAGTGATTGGCTTTGAAGTTGAA 780  
QY 781 GTGGGACGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTGC 840  
Db 781 GTGGGACGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTGC 840  
QY 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGCTGACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGCTGACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTTTCTGGCGCGCTATTTCGCAATT 960  
Db 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTTTCTGGCGCGCTATTTCGCAATT 960  
QY 961 CATCATGCCGCAACAGCCGCAACTGATGAGTATATCGTGCTCGCGGTCAACCACTATTGT 1020  
Db 961 CATCATGCCGCAACAGCCGCAACTGATGAGTATATCGTGCTCGCGGTCAACCACTATTGT 1020  
QY 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCCCTGCTCAACGGATTGCTTATGGAT 1080  
Db 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCCCTGCTCAACGGATTGCTTATGGAT 1080  
QY 1081 TAAAGGACCAAGCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAGCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGT 1140  
QY 1141 GGGAGCGCTGTAGCATCGCGAGGCGATCGGTGAAAATAATATGTCGATCGCGGTAAAC 1200  
Db 1141 GGGAGCGCTGTAGCATCGCGAGGCGATCGGTGAAAATAATATGTCGATCGCGGTAAAC 1200

RESULT 2  
US-09-711-164-274  
; Sequence 274, Application US/09711164  
; Patent No. 6589738  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF

FILE REFERENCE: ELITRA.008A  
; CURRENT APPLICATION NUMBER: US/09/711.164  
; PRIOR FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-9  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 274  
; LENGTH: 417  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(417)  
US-09-711-164-274  
Query Match 34.8%; Score 417; DB 4; Length 417;  
Best Local Similarity 100.0%; Pred. No. 2.9e-128; Indels 0; Gaps 0;  
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 758 GTGATTGCGTTTGAAGTGTGAAGTGGCAGCGCGGCTTACTTGAATTGGCTGGGAATC 817  
Db 1 GTGATTGCGTTTGAAGTGTGAAGTGGCAGCGCGGCTTACTTGAATTGGCTGGGAATC 60  
QY 818 CAGCAGTGGCGCGCGCTGTGCAATTGACCTTAAATCGCTGGCTTACTCAATCGCGT 877  
Db 61 CAGCAGTGGCGCGCGCTGTGCAATTGACCTTAAATCGCTGGCTTACTCAATCGCGT 120  
QY 878 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATCCCAAAAGTATTGTTT 937  
Db 121 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATCCCAAAAGTATTGTTT 180  
QY 938 CTGGCGCGCTATTTCGCAATTCATCATGCGCGCAACAGCCGCACTGATGCGATATATC 997  
Db 181 CTGGCGCGCTATTTCGCAATTCATCATGCGCGCAACAGCCGCACTGATGCGATATATC 240  
QY 998 GTGCTCGCGCTCACCACCTATTGTCGATATTATTGTCGATGATCGGTTACGCCACCTT 1057  
Db 241 GTGCTCGCGCTCACCACCTATTGTCGATATTATTGTCGATGATCGGTTACGCCACCTT 300  
QY 1058 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 1117  
Db 301 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360  
QY 1118 TTGCGCTCGTTTGTATGCTGGGAGCGCTGTTAGCATCGCGGAGCATGCGTGA 1174  
Db 361 TTGCGCTCGTTTGTATGCTGGGAGCGCTGTTAGCATCGCGGAGCATGCGTGA 417  
RESULT 3  
US-09-489-039A-895  
; Sequence 895, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489.039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 895  
; LENGTH: 645  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-895  
Query Match 32.6%; Score 391.8; DB 4; Length 645;  
Best Local Similarity 77.5%; Pred. No. 8.9e-120; Indels 3; Gaps 2;  
Matches 500; Conservative 0; Mismatches 142; Indels 3; Gaps 2;

QY 533 AAATAGACACACCGGAGTTTCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACA 592  
Dd |||||  
1 AAACAGACAATCAGGAGTTTCTCATGACCATGAGTGGTGGTTCGCTACCTGCTGACA 60  
QY 593 TCGATCATTTTAAACGCTGTGCGCAGGCTCTGCTGCAATCAACACTATGACCACTCGCTC 652  
Dd |||||  
61 TCAATATTCTCAGCTGTGCGCGGTTCCGAGCGATTATACCATGACCACTCCATT 120  
QY 653 AACCCAGGTTA--TCGCGCGGTGGCTCTATTGCTGGCTTCAGACCGGACTGGCGATT 710  
Dd |||||  
121 AACCCAGGATACCGCGCGCGCGCTGATTGCGGTTTTCAGACTGGGCTGGCCATT 180  
QY 711 CATATTGTGCTGGTGGCG-TGGGTTGGGACGCTATTTCGCGCTCAGTGATTGCGTTT 769  
Dd |||||  
181 CATATCGTGGTGGCTTGGTCTCGGCACTCTCTTCGCGCTCGGCTGGCCATT 240  
QY 770 GAAGTGTGAAGTGGGAGGCGCGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGC 829  
Dd |||||  
241 GAGGTGCTGAAATGGGCGCGCGCGCTACCTGATCTGGCTGGGTATCCAGCAATGGCGG 300  
QY 830 GCCGTGGTGCAATGACCTTAATCGCTGGCTCTACTCAATCGCTGACATTTGTTTC 889  
Dd |||||  
301 GCGGAGGGGCGATCGACCTGAATACGCTGGCGAAGGCGACGCGCGGCAAACTGTTT 360  
QY 890 CAGCGCGCAGTTTGTGTAATCTCACCATAATCCCAAAAGTATGTGTTTCTGGCGGCTA 949  
Dd |||||  
361 CAGCGCGCGTGTTCGTAATCTCACCATAATCCCAAAAGCATGTGTTTCTGGCGCGCTG 420  
QY 950 TTTCCGCAATTCATGATCGCGCAACAGCCGCACTGATGAGTATATCGTCTCGGCGTC 1009  
Dd |||||  
421 TTTCCGAGTTTATCTGCGCGCACCGCGAGGTGATGCAATATATCGTCTGGGCGTC 480  
QY 1010 ACACATATTGCTGATATTATTGATGATCGGTTACGCCACCTTGTCTCAACGGATT 1069  
Dd |||||  
481 ACCACATCGTCTGATATCATTTGATGATGCTATGCGACCTTGGCGCAGCGATT 540  
QY 1070 GCTCTATGATTAAAGGACCAAGCAGATGAGCGCTGATAAGATTTCGGCTCGTTG 1129  
Dd |||||  
541 TCCGCAATGATCAAGGGCCAAAGCAGATGAGCGCTTGACAAAGTCTTGTTCACCTG 600  
QY 1130 TTTATGCTGGTGGGAGCGCTGTAGCATCGCGAGGCGATGGGTGA 1174  
Dd |||||  
601 TTTATGCTGGTGGGAGCGCTGTCTCGCTCCGCGCCATGCGCTGA 645

RESULT 4

US-09-489-039A-851/c  
; Sequence 851, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 851  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-851

Query Match 24.8%; Score 297.2; DB 4; Length 1041;  
Best Local Similarity 75.7%; Pred. No. 3,1e-88;  
Matches 368; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGCGCAGTATATAGCGTTTACGCCAC 60  
Dd |||||  
494 AGAAAAGCGTGGCAATCGCTCCGCCCATAGATGAGCAAGGATAAACCGTTTTCGCCAG 435

QY 61 GGACCGGCTGAACCTCTCTGCTGCCAGAAATCCGCGCAGATCATCAACATAATCAATAAG 120  
Dd |||||  
434 TGTCCAGGAACCACTCTCTGCTGCCAGAGCGCGCGCAGGTATCAACATAGTCGCTGAAG 375  
QY 121 CGATTACATGCCCGGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTCCCTGCCCGCA 180  
Dd |||||  
374 TTAACCACTGTCCGCGTGGTATCTGAGAGCAGCGCGCGCGGCTCCCTGCCCGCGA 315  
QY 181 TGGTCGATGATTAAGACATCAACCCCAAAATGGAACAGGTCAATAGGCCAGTTCCGCATAT 240  
Dd |||||  
314 TGGTCAATGATCATCACTCAAGCCGCGAGTGAAGAGAGATCGTAGGCCACTTCAGCATAC 255  
QY 241 TTTACGTAGCTCTCAATACGCCCGGCGAGATGACTACACCCGCTCATGGTGTGTCG 300  
Dd |||||  
254 TTCACGTAGCTTTCAATGCGCGCGCGGCAAAATCAACACAGACGATCGTTATGTTGCGCG 195  
QY 301 CGAAAACGGACAAAGCGCACCGGAATGTCTATCCACACAGTAAACTCTGCTTCATCACGC 360  
Dd |||||  
194 CAAAAGCGGACAAAGCGTACCGGTACATCGTCGACGCTTTAAATTCAGCTTCTTCGCGC 135  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTTCTCTGTT 420  
Dd |||||  
134 TGTGCCAGAAATCCGTCAGCGGCCCATGTAAAGCGCGCAACCGGTTTTCTCTGTT 75  
QY 421 TCCAGTCTTTTGTGCTGAAACATCGGGTAAATTCGCTCTTTAAACACAGTAAATCGT 480  
Dd |||||  
74 TCCAGTCTTTTCTGCCCCAAACAATCGGGTTTACACCCCTGTTAAACCGCTGAAATCGT 15  
QY 481 TTTTTT 486  
Dd |||||  
14 TTTTTT 9

RESULT 5

US-09-711-164-107/c  
; Sequence 107, Application US/097111164  
; Patent No. 6589738  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERE  
; FILE REFERENCE: ELITRA.008A  
; CURRENT APPLICATION NUMBER: US/09/711,164  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-9  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107  
; LENGTH: 263  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-711-164-107

Query Match 17.3%; Score 207.4; DB 4; Length 263;  
Best Local Similarity 94.3%; Pred. No. 8.8e-59;  
Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 861 CCTCTACTCAATCGCGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACCAATC 920  
Dd |||||  
263 CCCCTACTCAATCGCGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACCAATC 204  
QY 921 CCAAAGTATTGTTTCTGCGCGGCTATTTCGCAATTCATCATGCCGCAACAGCCGC 980  
Dd |||||  
203 CCAAAGTATTGTTTCTGCGCGGCTATTTCGCAATTCATCATGCCGCAACAGCCGC 144  
QY 981 AACTGATGCAGTATATCGTCTC-GGCGTCACCACTATTGTTGTCGATATTATTGATG 1039  
Dd |||||  
143 AACTGATGCAGTATATCGTCTCGGCGTCACCACTATTGTTGTCGATATTATTGATG 84  
QY 1040 ATC-GGTTACGCCACCTTCTCAACGGATTG--CTCTATGGATTAAAGGACCAAGCAG 1096  
Dd |||||

D<sub>b</sub> 83 ATCGGGTTACGCCACCCCTTGCTCTCAACGGGATGGCTCTAAGGGATTAAAGGCCCAACCAA 24

Q<sub>y</sub> 1097 ATCAAGGCGCTGAATAAGATTTT 1119  
|||||

D<sub>b</sub> 23 ATCAAGGCGCTGAATAAGATTTT 1  
|||||

## RESULT 6

```

US-09-543-681A-238/c
; Sequence 238, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO
; TITLE OF INVENTION: DIAGNOSTICS AND THERA
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 238
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-238

```

	Query Match	12.1%;	Score 144.8;	DB 4;	Length 1038;
	Best Local Similarity	59.5%;	Pred. No. 1.2e-37;		
	Matches 245;	Conservative 0;	Mismatches 167;	Indels 0;	Gaps 0;
Qy	14	AGATCGCACCGCCCATCGAATGTGTCAGTATATATAGCGTTACGCCACGGACCGGCTGAA	73		
Db	472	AAATAGCGCGCCCATCGAATGGGCTAACGCCAAATAGTCTGATAGTATAGCAAAA	413		
Qy	74	CCTCCTGCTGCCAGAAATGCCGCCAGATCATCAACATATCATTAACGGATTAAATGCC	133		
Db	412	CGATAGTATCAATAAATGTCGATAAATCATCGACATATCACTGAATTTTCTACGTGCC	353		
Qy	134	CGAGATCGGGATCGGCTAACAGGCGACCGGGAACGTCCCTGCCCCCGGATGTCGATGATTA	193		
Db	352	CTTTTTCGGATCATCTAATAAGCGCCCTGATAATCCCTTGGCCCCCGATGATCGAGAAGAA	293		
Qy	194	AGACATCAAAACCCCAAATGGAACAGGTCATAGGCCAGTTCGGCATATTTTACGTAGCTCT	253		
Db	292	AAATATCATAAACCAAGGTGATAAAAATCGAAGGCCACITTCAGGGTATTTTACATAACTTT	233		
Qy	254	CAATACGCCCCCGGCAGATGACTACCACCCGGTCATGGTGTGTGCGGAGAAAACGGACAA	313		
Db	232	CACACGGCCAGAGCAGATCACCACCGCTTTTGTGTGATGAAGAGCGCCAATGAACAT	173		
Qy	314	AGCGCACCGGAATGTATCCACACCCAGTAAACTCTGCTTCATCACGCTGACGCCAGAAAT	373		
Db	172	AGCGAATGCTCACGCCATCGACTCCGGTAAATTCGCTTCCTGCTGTTGGTTCCAAAAGT	113		
Qy	374	CAGTCAGCGGTCCCATGGTAAAAAGCAGCAAAACCGCGTTTTCTCTTTTCCCA	425		
Db	112	CCATFAAGATACCATTCGTAAAGCAGAGAAATGCTTTTTCACGAGATAACCA	61		

## RESULT 7

```

US-09-252-991A-6447
; Sequence 6447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: AERUGINOSA FOR DI
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,7
; PRIOR FILING DATE: 1998-02-18

```

; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6447  
 ; LENGTH: 834  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6447

Query Match 9.4%; Score 112.4; DB 4; Length 834;  
 Best Local Similarity 53.5%; Pred. No. 6.3e-27;  
 Matches 324; Conservative 0; Mismatches 276; Indels 6; Gaps 4;

Qy 572 TGGTTTGGCTACCTGCTGACATCGATCATTTTAAAGCTGTGCGCAGGCTCTGGTGCAATC 631  
 Db 220 TGGTTCGCTCTTCTTCGCTGCTGGGCCATCAGCCTGTGCGCGGGCGCGGAGCCATC 279  
 Qy 632 AACACTATGACCACCTCGCTCAACACAGGTTATCGCGCCGGTGGC--GTCATTGCTGGG 689  
 Db 280 GCCTCGATGCTCGGGTTGCAGTACGGCTTTGCGCGAGGCTACTGGAACGCGCTGGGC 339  
 Qy 690 CTTCAGACCGGACTGGCGATTTCATATTGCTGCTGGTTGGCG-TGGGGTTGGAGCAGCTATT 748  
 Db 340 CTGCAGATCGGCTTGGCCCTGCAGATCGCCATTGTCGCCCGCGCTCGGTGCGTTGCTG 399  
 Qy 749 TCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCAGGCGCGGCTTACTTGATTGG 808  
 Db 400 GCGACCTCGGCACTGGCTTTCAGCCTGATCAAGTGGTTCGGCGTGGCCTACCTGGTGAC 459  
 Qy 809 CTGGGAATCCAGCAGTGGC--GCGCGCTGGTGCATTTGACCTTAAATCGSTGGCCTCTA 866  
 Db 460 CTGGCGGTGCGCCAGTGGCAGGCGCGCCACAGGCTTGAGCACCGATGCGGAACGGCCT 519  
 Qy 867 CTCGAATCGCTCGACATT-TGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATCCCAA 925  
 Db 520 CTGGGGCGACCGTTGACCTGGTGGTGGTGGTTCCTGGTCAACGCCAGCAATCCCAAG 579  
 Qy 926 AGTATTGTTTCTGGCGCGCTATTTCGCAATTTCATGCGCGCAACAGCGCAACTG 985  
 Db 580 GCGGTGATCTTCATGTCGCGGTGCTGCGCAGTTCATGACCGCACCCGCGCTGCTG 639  
 Qy 986 ATGCAGTATATCGTCTCGCGCGTCAACACTATTGTGGTGCATATTATTGTGATGATCGGT 1045  
 Db 640 GCGCAATACCTGATCATGCGCGGCAACCATGATGCTGCTGCAGCTGATGTCATGGCGGC 699  
 Qy 1046 TAGCCACCTTGTCTAACGGATTGCTCTATGGAATTAAGGACCAAGCAGATGAAGCG 1105  
 Db 700 TACACCGGGTGGCTGCGCGCGTGTACGCGTACTGCGTTCGCGCGCCAGCAGAAGCTG 759  
 Qy 1106 CTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTAGCATCGCGGAG 1165  
 Db 760 GTGAACCGTACCTTCGCCAGCCTGTTGCTCGGTGCGCGGGGTGCTGCGCAGGTACGC 819  
 Qy 1166 CATGCG 1171  
 Db 820 CGAGCG 825

RESULT 8  
 US-09-252-991A-6103/c  
 ; Sequence 6103, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6103  
LENGTH: 930  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6103

Query Match 8.5%; Score 102.4; DB 4; Length 930;  
Best Local Similarity 55.6%; Pred. No. 1.4e-23;  
Matches 278; Conservative 0; Mismatches 216; Indels 6; Gaps 4;

QY 572 TGGTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCCAGGCTCTGGTGCAATC 631  
DB 504 TGGTTGCTCTTCTTCTCGCTGTGGCCATCAGCCTGTGCCCGGCGCGAGCCATC 445  
QY 632 AACATATGACCACTCGCTCAACACAGGTTATCCGGCCGGTGGC--GTCTATTGCTGGG 689  
DB 444 GCCTCGATGCTCCTGGGTTGAGTACGGCTTTGCGCGAGGCTACTGGAACGCGTGGGC 385  
QY 690 CTTACAGACGGACTGGCGATTCTATATTGTGCTGGTGGCG-TGGGTTGGACGCTATT 748  
DB 384 CTGCAGATCGGCTTGGCCCTGCAGATCGCCATTTGTCGCCCGCGCTGGTGGCTG 325  
QY 749 TCCCGCTCAGTATTGCGTTTGAAGTGGTGAAGTGGGCGAGGCGGCTTACTTGAATTGG 808  
DB 324 GCGACCTCGGCACTGGCTTTAGCTGATCAAGTGGTTCCGGGTGGCTTACCTGGTGATC 265  
QY 809 CTGGGAATCCAGCAGTGGC--GCGCGCTGGTGAATGACCTTAATCGTGGCTCTA 866  
DB 264 CTGGCGTGGCCAGTGGCAGGCGCGCCACAGGCTTGAGACCGATGGGAAACGGCCT 205  
QY 867 CTCATCG-CGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATCCCAA 925  
DB 204 CTGGGCGGACCGTTGACCTGTTGCTGGTGGTGTCTCTGTTCAACGCCAGCAATCCCA 145  
QY 926 AGTATTGTTTCTGGGCGGCTATTTCGCAATTCAATCATGCGCGCAACAGCGCAACTG 985  
DB 144 GCGGTGATCTTCATGCTCGCGGTGCTGCGCAGTTCATCGACCGCACCGCGCTGCTG 85  
QY 986 ATGAGTATATCGTGTGCGGCTGACCACTATTGTGTCGATATTATTGTGATGATCGGT 1045  
DB 84 GCGCAATACCTGATCATGGCGCGCACCATGATCGTGTGACCTGATCGTATGGCGCGC 25  
QY 1046 TACGCCACCTTGTCTCAACG 1065  
DB 24 TACACCGGCTGGCTGGCG 5

RESULT 9  
US-09-328-352-777  
Sequence 777, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 777  
LENGTH: 663  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-777

Query Match 6.4%; Score 76.6; DB 4; Length 663;  
Best Local Similarity 49.1%; Pred. No. 4.2e-15;  
Matches 289; Conservative 0; Mismatches 294; Indels 6; Gaps 3;

QY 551 TTCATCATGACCTTAGAATGGTGGTTGCTTACCTACCTGCTGACATCGATCATTTTAACGCTG 610  
DB 31 TTCGCTATGCTTTTGCAAGTTTGGTTTGTCTTATATGTTGGCTGTGGTTCATTAGTATT 90

QY 611 TCGCCAGGCTCTGGTGAATCAACTATGACCACTCGCTCAACCAAGTTATCCGGCC 670  
DB 91 TCTCCAGGAGCAGGGCAATTGCTTCTATGTCGAGCGCTTAAATTATGGCTTTAGACAT 150  
QY 671 GGTGGC--GTCTATTGCTGGCTTCAAGCCGACTGGCGATTCAATATTGCTGCTGGC 728  
DB 151 GGCTACTGGAATGCAATTGGTTTGCAATCGCTTGTAAATTCAAATCATGATTGTAGCG 210  
QY 729 GTGGG-GTGGGACGCTATTTTCCGCTCAGTATGTCGTTTGAAGTGTGAAGTGGGCA 787  
DB 211 GCGGTTGAGTGTCTATTCGCAACGACACCATTAGCCTTTTCAAGCGTTAAGTGGTTT 270  
QY 788 GCGCGGCTTACTTGAATTGGCTGGGAATCCAGACTGGCGCGCTGGTGA--ATT 844  
DB 271 GGGGTAGCTTATTATTGATTAGCCTATTGCGAGTGGACAGCCTGTAAAGATATA 330  
QY 845 GACCTTAATCGCTGGCTCTACTCAATCGGCTCGACATTTGTTCCAGCGCGCTTTT 904  
DB 331 GAAATTCACACGAAAGGATAAATCTGTTTCTGCTCTGCTGTTCAATGGATTGTA 390  
QY 905 GTGAATCTCACCAATCCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTCATC 964  
DB 391 GTCAATATCAGTAAACCTAAAGCCATCGTATTTTATTGGCAGTTTGCCTCAGTTCTTA 450  
QY 965 ATGCCGCAACAGCCGCACTGATGATGATGATATATCGTGTGGCGCTGACCACTATTGTC 1024  
DB 451 GATTTAAGCAACCTCAATGGATACATATCTGATTATGGCAGCCACTATGGTTACGATT 510  
QY 1025 GATATTATTGATGATCGGTTTACGCCACCTTGTCTCAACGATTGCTCTATGATTA 1084  
DB 511 GATTGATTGTAATGGCTGTATACAGGGCTAGCTTCAAAAGTTTAAAGATTGCTACGT 570  
QY 1085 GGACCAAGCAGATGAAGCGCTGAATAAGATTTCGCTCGTTGTTTA 1133  
DB 571 TCTCCTAAGCAACAAATATTAAACCGTGGTTTTCAGTTCATGTTTA 619

RESULT 10  
US-09-489-039A-2622  
Sequence 2622, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 2622  
LENGTH: 645  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2622

Query Match 4.9%; Score 59; DB 4; Length 645;  
Best Local Similarity 46.1%; Pred. No. 2.9e-09;  
Matches 197; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 730 TGGGGTTGGACGCTATTTTCCGCTCAGTGAATGAGTGTGAAGTGGGAGG 789  
DB 204 TGGTTAGTGGCTACTGGCGGTTTCTGAGTTGGCTTATACCTTATGAAATGGTGGG 263  
QY 790 CGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATGACCT 849  
DB 264 AGCAGGATACCTTTGCTGGCTGGGATCCAGTTGCTCTCGACCCAGCCAGCAATTAA 323  
QY 850 TAAATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGAGTTTGTGAA 909  
DB 324 CACCCACCGCGCGAATCAGATTCCACAGTAACCTGTTCTGCTGGGATGCTGGGAA 383

QY 910 TCTCACCATCCCAAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCC 969  
Db 384 TGTGCTTAACCCCAAAATGGCGCTCTCTACGTTTCGTTTTTACCTCAGTTTATCCCGC 443  
QY 970 GCAACAGCGGCAACTGATGCAGTATATCGTGTCTGGCGTCAACACTATTGTGGTCGATAT 1029  
Db 444 CGGGCAATCAACCGTGAGCTGGACTTTCTCCTGTGCTATCCATGATTAATTGGTAC 503  
QY 1030 TATTGTGATGATCGGTACGCCACCCCTGCTCAACGGATTGCTCTATGGATTAAAGGACC 1089  
Db 504 GCTGTGGTGTAAACGCTGATTACCGCCACTCGCTATGCCGAGGCATCCTGAAAAACC 563  
QY 1090 AAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGTTGGAGCGCT 1149  
Db 564 AGCCGTAGTGAATGGATGGATCGGACCGCGCTGTTCTGTTTATTGCTGCGGAA 623  
QY 1150 GTTAGCA 1156  
Db 624 GCTGGCA 630

RESULT 11  
US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: P8186P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Query Match 3.8%; Score 46; DB 4; Length 1830121;  
Best Local Similarity 52.5%; Pred. No. 0.0077;  
Matches 128; Conservative 0; Mismatches 110; Indels 6; Gaps 1;  
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60

Db 689153 AAATAATAGGTTGCAATCAATCGCCCATTTGAATGAGCAAGCAAAATGCTGCGTTGAATAG 689212  
QY 61 GGACCGGGCTGAACCTCTGCTGCCAGAATCCCGCAGATCATCAACATAATCAATAAG 120  
Db 689213 CTGAAAAGTGGGTTACTTTTCGATGATTTTCGCCATATCATCGACATAGAAACGAAAT 689272  
QY 121 CGATTAACTGCCGAGATCGGATCGGCTAACAGGCGACCGGAACGTCCTGCCCGCGA 180  
Db 689273 TCATCTAAATGCCCTTTTGTAGGA-----ATAATAGCTGTGAATAGCCTTGACCGCGA 689326  
QY 181 TGGTCGATGATTAAGACATCAACCCCAAAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 689327 TGATCAATAGCAATACATCGTAACCTTGATGATATAAAATCATAGCGAGCTCTGACCAT 689386  
QY 241 TTTA 244  
Db 689387 TTCA 689390

Query Match 3.8%; Score 46; DB 4; Length 1830121;  
Best Local Similarity 52.5%; Pred. No. 0.0077;  
Matches 128; Conservative 0; Mismatches 110; Indels 6; Gaps 1;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 16:45:05 ; Search time 509.752 Seconds  
(without alignments)  
8764.284 Million cell updates/sec

Title: US-09-847-392-1

Perfect score: 1200

Sequence: 1 agaaataatgtgagatgcg.....atgtcgatgcggcgtaaac 1200

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1200	100.0	1200	9	US-09-927-395-1
2	1200	100.0	1200	9	US-09-847-392-1
C 3	446	37.2	1023	12	US-10-282-122A-20578
C 4	446	37.2	1023	15	US-10-369-493-47314
5	417	34.8	417	12	US-10-282-122A-6496
6	417	34.8	417	14	US-10-287-274-274
7	382.2	31.9	621	12	US-10-282-122A-39598
8	375	31.2	618	12	US-10-282-122A-23973
9	357.4	29.8	618	12	US-10-282-122A-20027
10	348.2	29.0	623	12	US-10-282-122A-36679
C 11	298.8	24.9	1017	12	US-10-282-122A-7563
C 12	297.2	24.8	1017	12	US-10-282-122A-39597
C 13	290.8	24.2	990	12	US-10-282-122A-20024
C 14	276.4	23.0	990	12	US-10-282-122A-23738
15	275	22.9	621	12	US-10-282-122A-41840

C 16	262.8	21.9	1026	12	US-10-282-122A-36678	Sequence 36678, A
C 17	259	21.6	259	12	US-10-282-122A-1269	Sequence 1269, Ap
C 18	207.4	17.3	263	14	US-10-287-274-107	Sequence 107, App
C 19	181.8	15.2	1014	12	US-10-282-122A-41838	Sequence 41838, A
C 20	165.8	13.8	975	15	US-10-369-493-24057	Sequence 24057, A
C 21	156.2	13.0	990	15	US-10-369-493-44919	Sequence 44919, A
C 22	149.8	12.5	618	12	US-10-282-122A-40780	Sequence 40780, A
C 23	144.8	12.1	1002	12	US-10-282-122A-32814	Sequence 32814, A
C 24	112.4	9.4	630	12	US-10-282-122A-30630	Sequence 30630, A
C 25	104	8.7	630	12	US-10-282-122A-33154	Sequence 33154, A
C 26	86.8	7.2	633	12	US-10-282-122A-31699	Sequence 31699, A
C 27	80	6.7	627	12	US-10-282-122A-14918	Sequence 14918, A
C 28	76	6.3	1017	12	US-10-282-122A-40779	Sequence 40779, A
C 29	70.8	5.9	978	15	US-10-369-493-35643	Sequence 35643, A
C 30	65.6	5.5	954	12	US-10-282-122A-31172	Sequence 31172, A
C 31	64.6	5.4	543	12	US-10-282-122A-8689	Sequence 8689, Ap
C 32	62.4	5.2	864	15	US-10-369-493-35283	Sequence 35283, A
C 33	62.4	5.2	864	15	US-10-369-493-38570	Sequence 38570, A
C 34	62.4	5.2	882	15	US-10-369-493-37975	Sequence 37975, A
C 35	62.4	5.2	882	15	US-10-369-493-38253	Sequence 38253, A
C 36	60.8	5.1	795	15	US-10-369-493-31514	Sequence 31514, A
C 37	58.8	4.9	633	12	US-10-282-122A-11642	Sequence 11642, A
C 38	58.2	4.9	792	15	US-10-369-493-45471	Sequence 45471, A
C 39	58.2	4.9	822	15	US-10-369-493-44326	Sequence 44326, A
C 40	53.6	4.5	1092	12	US-10-282-122A-15543	Sequence 15543, A
C 41	46	3.8	942	12	US-10-282-122A-22067	Sequence 22067, A
C 42	46	3.8	1830121	14	US-10-329-960-1	Sequence 1, Appli
C 43	46	3.8	1830121	15	US-10-329-670-1	Sequence 1, Appli
C 44	45.6	3.8	810	15	US-10-369-493-34253	Sequence 34253, A
C 45	43.2	3.6	536	16	US-10-338-110-119	Sequence 119, App

ALIGNMENTS

RESULT 1

US-09-927-395-1  
; Sequence 1, Application US/09927395  
; Patent No. US20020058314A1  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENIAMTOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD  
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/927,395  
; CURRENT FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 09/396,357  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: RU98118425  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (557)..(1171)  
US-09-927-395-1

Query Match 100.0%; Score 1200; DB 9; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATGCGACCGCCCATCGAATGCGCCAGTATATAGCGTTTACGCCAC 60

Db 1 AGAAATAATGTGGAGATGCGACCGCCCATCGAATGCGCCAGTATATAGCGTTTACGCCAC 60



Db 481 TTTTTCAGCGTGCCTGACACAAACGCTGCGAGTAGCGTATTGTGGCACAATAATAGAC 540  
QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
Db 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAT 600  
QY 601 TTTAACGCTGTGCCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGG 660  
Db 601 TTTAACGCTGTGCCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGG 660  
QY 661 TTATCCGGCGGTGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTGC 720  
Db 661 TTATCCGGCGGTGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTGC 720  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTTGGGTTTGAAGTTGAA 780  
Db 721 TGGTTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGATTTGGGTTTGAAGTTGAA 780  
QY 781 GTGGCAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCCGCTGGTGC 840  
Db 781 GTGGCAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCCGCTGGTGC 840  
QY 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 900  
Db 841 AATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 900  
QY 901 TTTTGTGAATCTCACCACATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATT 960  
Db 901 TTTTGTGAATCTCACCACATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATT 960  
QY 961 CATCATGCCGCAACAGCGCAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Db 961 CATCATGCCGCAACAGCGCAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
QY 1021 GGTGATATTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
Db 1021 GGTGATATTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 1081 TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
Db 1081 TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT 1140  
QY 1141 GGGAGCGCTGTAGCATCGCGGAGGATCGGTGAAAATAATGTCGGATCGCGGTAAC 1200  
Db 1141 GGGAGCGCTGTAGCATCGCGGAGGATCGGTGAAAATAATGTCGGATCGCGGTAAC 1200

RESULT 3

US-10-282-122A-20578/c  
; Sequence 20578, Application US/10282122A  
; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282.122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20578  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-282-122A-20578

Query Match 37.2%; Score 446; DB 12; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 2.4e-144;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db 446 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 387  
QY 61 GGACCGGGCTGAACCTCCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAAG 120  
Db 386 GGACCGGGCTGAACCTCCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAAG 327  
QY 121 CGATTAAACATGCCCGAGATCGGATCGGTAACAGGCGACCGGAACGTCCTGCCCGGA 180  
Db 326 CGATTAAACATGCCCGAGATCGGATCGGTAACAGGCGACCGGAACGTCCTGCCCGGA 267  
QY 181 TGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 240  
Db 266 TGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTATAGGCCAGTTCCGCATAT 207  
QY 241 TTTACGTAGTCTCAATACGCCCGGCGAGATGACTACACCCGGTATGGTGTGTGCG 300  
Db 206 TTTACGTAGTCTCAATACGCCCGGCGAGATGACTACACCCGGTATGGTGTGTGCG 147  
QY 301 CGAAAAACGGACAAAGCGCACCGGAATGTATCCACACAGTAAACTCTGCTTCATCAGC 360  
Db 146 CGAAAAACGGACAAAGCGCACCGGAATGTATCCACACAGTAAACTCTGCTTCATCAGC 87  
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGAGCAGCAACCGGTTTTCTTTGTT 420  
Db 86 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGAGCAGCAACCGGTTTTCTTTGTT 27  
QY 421 TCCAGTCTTTTGTGCTGCTGAAACAT 446  
Db 26 TCCAGTCTTTTGTGCTGCTGAAACAT 1

RESULT 4

US-10-369-493-47314/c  
; Sequence 47314, Application US/10369493  
; Publication No. US20030233675A1

GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 47314
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-369-493-47314

Query Match      37.2%; Score 446; DB 15; Length 1023;
Best Local Similarity 100.0%; Pred. No. 2.4e-144;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATATGTGGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 446 AGAAATATGTGGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 387

QY 61 GGACCGGCTGAACCTCTGCTGCCAGATGCCGCCAGATCATCAATAATCAATTAAG 120
Db 386 GGACCGGCTGAACCTCTGCTGCCAGATGCCGCCAGATCATCAATAATCAATTAAG 327

QY 121 CGATTAAACATGCCCGAGATCGGGATCGGCTAACAGGCGACCGGAACGTCCCTGCCCGCA 180
Db 326 CGATTAAACATGCCCGAGATCGGGATCGGCTAACAGGCGACCGGAACGTCCCTGCCCGCA 267

QY 181 TGTCTGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCTATAGCCAGTTCCGCATAT 240
Db 266 TGTCTGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCTATAGCCAGTTCCGCATAT 207

QY 241 TTTACGTAGCTCTCAATACGCCCCGGGCGAGATGACTACCAACCGGTCTATGTTGTTGG 300
Db 206 TTTACGTAGCTCTCAATACGCCCCGGGCGAGATGACTACCAACCGGTCTATGTTGTTGG 147

QY 301 CGAAACCGGACAAAGCGACCGGAATGTCATCCACACAGTAAACTCTGTTTCAATCAGC 360
Db 146 CGAAACCGGACAAAGCGACCGGAATGTCATCCACACAGTAAACTCTGTTTCAATCAGC 87

QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 420
Db 86 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAACCGGTTTCTCTTGT 27

QY 421 TCCAGTCTTTTGTCTGCTGAAACAT 446
Db 26 TCCAGTCTTTTGTCTGCTGAAACAT 1
```

```
RESULT 5
US-10-282-122A-6496
; Sequence 6496, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6496
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-6496

Query Match      34.8%; Score 417; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.8e-134;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 GTGATTGCGTTTGAAGTGTGAAGTGGGAGGCGCGGCTTAAATCGTGGCTTACTCAATCGGT 817
Db 1 GTGATTGCGTTTGAAGTGTGAAGTGGGAGGCGCGGCTTAAATCGTGGCTTACTCAATCGGT 60

QY 818 CAGCAGTGGGCGCGCGCTGTGTCGAATTGACCTTAAATCGTGGCTTACTCAATCGGT 877
Db 61 CAGCAGTGGGCGCGCGCTGTGTCGAATTGACCTTAAATCGTGGCTTACTCAATCGGT 120

QY 878 CGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTT 937
Db 121 CGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTT 180

QY 938 CTGCGGCGGTATTTCCGCAATTCATCATCGCGCAACAGCGCAACTGATGCGTATATC 997
Db 181 CTGCGGCGGTATTTCCGCAATTCATCATCGCGCAACAGCGCAACTGATGCGTATATC 240

QY 998 GTGCTCGGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCGGTACGCCCTT 1057
Db 241 GTGCTCGGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCGGTACGCCCTT 300

QY 1058 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGGTGAATAAGATT 1117
Db 301 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGGTGAATAAGATT 360

QY 1118 TTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTAGCATCGGCGAGGATCGGTGA 1174
Db 361 TTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTAGCATCGGCGAGGATCGGTGA 417

RESULT 6
US-10-287-274-274
; Sequence 274, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
```



;; PRIOR FILING DATE: 1999-11-09  
;; PRIOR APPLICATION NUMBER: US 09/711164  
;; PRIOR FILING DATE: 2000-11-09  
;; NUMBER OF SEQ ID NOS: 469  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 274  
;; LENGTH: 417  
;; TYPE: DNA  
;; ORGANISM: Escherichia coli  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(417)  
US-10-287-274-274

Query Match 34.8%; Score 417; DB 14; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.8e-134;  
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 758 GTGATTGGCTTTGAAGTGTGAAGTGGCGAGCGCGCTTACTTGATTGGTGGGAATC 817  
Db |||||  
1 GTGATTGGCTTTGAAGTGTGAAGTGGCGAGCGCGCTTACTTGATTGGTGGGAATC 60  
QY 818 CAGCAGTGGCGCGCGCTGGTGAATGACCTTAATCGTGGCTCTACTCAATCGCGT 877  
Db |||||  
61 CAGCAGTGGCGCGCGCTGGTGAATGACCTTAATCGTGGCTCTACTCAATCGCGT 120  
QY 878 CGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAAGATTGTGTTT 937  
Db |||||  
121 CGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAAGATTGTGTTT 180  
QY 938 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGAGTATATC 997  
Db |||||  
181 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGCAACTGATGAGTATATC 240  
QY 998 GTGCTCGGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGCCACCCCTT 1057  
Db |||||  
241 GTGCTCGGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGCCACCCCTT 300  
QY 1058 GCTCAACGCGATTGCTCTATGGATTAAAGGACCAAGCAGATGAGGCGCTGAATAAGATT 1117  
Db |||||  
301 GCTCAACGCGATTGCTCTATGGATTAAAGGACCAAGCAGATGAGGCGCTGAATAAGATT 360  
QY 1118 TTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTAGCATCGCGAGGCGATCGGTGA 1174  
Db |||||  
361 TTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTAGCATCGCGAGGCGATCGGTGA 417

RESULT 7  
US-10-282-122A-39598  
; Sequence 39598, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 39598  
;; LENGTH: 621  
;; TYPE: DNA  
;; ORGANISM: Salmonella typhi  
US-10-282-122A-39598  
Query Match 31.9%; Score 382.2; DB 12; Length 621;  
Best Local Similarity 78.1%; Pred. No. 3.4e-122;  
Matches 485; Conservative 0; Mismatches 133; Indels 3; Gaps 2;  
QY 557 ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCATTTTAAACGCTGTGCCCA 616  
Db |||||  
1 ATGACCTTTGAATGGTGGTTGCCTATCTACTGACCTCAACTCTGTGAGTCTTTCTCCG 60  
QY 617 GCCTCTGTGCAATCAACACTATGACACCTCGCTCAACACCGTTATCCGGCCG--GTG 674  
Db |||||  
61 GGTTCAGGCGCCATCAATACCATGACGACGTCTATCAACCATGGATATCGTGGCGCAGCG 120  
QY 675 GCGTCTATTGCTGGCTTCAGACCGGACTGCGGATTCATATTTGCTGTTGGGCTGGGG 734  
Db |||||  
121 GCTTCTATCGCGGGCTCCAGACCGGCTGGGATACATATCGTACTGGTGGGCTGGGA 180  
QY 735 TTGGG-ACGCTATTTTCCGCTCAGTGAATTCGTTTGAAGTGTGAACTGGGCGAGCGCG 793  
Db |||||  
181 CTGGGTACGCTCTTTTCGCGCTCGCTCATCGCTTTTGAATTTCTGAAATGGGCTGGCGCG 240  
QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATTTGACCTTAAA 853  
Db |||||  
241 GCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCTATCGATCTGCAT 300  
QY 854 TCGCTGGCTCTACTCAATCGCTCGACATTTTGTCCAGCGCGCAGTTTGTGATCTC 913  
Db |||||  
301 ACTCTGCCCGCAGACGCAATCGCGGGTGGCTGTTTCAACGCGGATATTTGTCAATCTA 360  
QY 914 ACCAATCCCAAAAGTATTGTTTCTGCGGGCGCTATTTCGCAATTCATCATGCCGCAA 973  
Db |||||  
361 ACCAATCCCAAAAGTATTGTTTCTTCCGCGCTGTTCCGCAATTCATCATGCCGCGAG 420  
QY 974 CAGCCGCAACTGATGAGTATATCGTCTCGGCTGCGGCTCACCACCTATTGTTGGTGCATATTATT 1033  
Db |||||  
421 CAACCGCAACTGGCGCAGTACCTCATTTCTGGCGTCCACCAAGATTGTTGGTGCATATTATT 480  
QY 1034 GTGATGATCGGTTACGCCACCCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAAAG 1093  
Db |||||  
481 GTGATGACCGGTTACGCCACCGCTGGCGCAGCGCATTTGCCCTGGATTAAAGGACCAAAG 540  
QY 1094 CAGATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA 1153  
Db |||||  
541 CAGATGAAGCGCTGAATAAGCGTTTGGTTGTTTATGCTGGTGGGCGCTGCTG 600  
QY 1154 GCATCGGCGAGGCGATCGGTGA 1174  
Db |||||  
601 GCGTCGGCAAGACACGCGTGA 621

RESULT 8  
US-10-282-122A-23973  
; Sequence 23973, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23973  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-10-282-122A-23973  
Query Match 31.2%; Score 375; DB 12; Length 618;  
Best Local Similarity 77.6%; Pred. No. 1.1e-119;  
Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;  
QY 557 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATGATCAATTTAAACGCTGCGCCA 616  
Db 1 ATGACCATTTGAGTGGTGGTTGCTACCTGCTGACATCAATTTAACTCAGCCTGCGCCG 60  
QY 617 GGCTCTGGTGCATCAACACATGATGACCACTCGCTCAACACCGGTTA--TCCGGCGCGTG 674  
Db 61 GGTTCCGGAGCGAATTAATACCATGACCACTCCATTAAACACGATACCGCGCGCGCG 120  
QY 675 GCGTCTATTGCTGGGCTTACAGCCGAGTGGCGATTCTATTGCTGGTGGCG-TGGG 733  
Db 121 GCGTCGATTGCGGCTTGCAGACCGGGCTGGTTATTCATATCGTGGTGGCGTTGGT 180  
QY 734 GTTGGGACGCTATTTCCCGCTCAGTGATGCGTTTGAAGTGTGAAGTGGGACGCGCG 793  
Db 181 CTCGGCACTCTCTCTCCCGCTGGTGGTGGCTTTGAGGTGCTGAATGGGCGCGCGCC 240  
QY 794 GCTTACTTGTGGTGGGAATCCAGAGTGGCGCGCTGGTGGTGAATGACCTTAA 853  
Db 241 GCCTACCTGATCTGGCTGGGTATCCAGCAATGGCGGGCGGCGAGGGCGGCGCTGAAT 300

QY 854 TCCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 913  
Db 301 ACGCTGGGAAGGCGCAGACGCGCGCAAACTGTTCCAGCGCGCGGTTCGTAATCTC 360  
QY 914 ACCAATCCAAAAGTATTGTTTCTGCGGCGCTATTTCGCAATTCATCATGCGCGAA 973  
Db 361 ACCAATCCAAAAGCATTTGTTTCTGCGGCGCCCTGTTCCGAGTTTATCTGCGCGAC 420  
QY 974 CAGCGCAACTGATGACATATATCTGTTCTGCGGCGTCACCACTATTGTTGTCGATATT 1033  
Db 421 CAGCGCAGGTGATGCAATATCTGTTCTGCGGCGCTGCGCATGCGTCGATATCAT 480  
QY 1034 GTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGAATAAGGACCAAG 1093  
Db 481 GTGATGATTGGCTATGCGACCCCTGGCGCAGCGCATTTCCGCATGGATCAAAAGGCGCAAG 540  
QY 1094 CAGATGAAGCGCTGAATAAGATTTTCGCTCGTTGTTTATGCTGTTGGAGCGCTGTTA 1153  
Db 541 CAGATGAAGCGCTGAACAAAGTCTTTGTTTCACTGTTTATGCTGTTGGGCGCTGCTC 600  
QY 1154 GCATCGGCGAGGCGCATGC 1170  
Db 601 GCCTCGCGCGCGCATGC 617  
RESULT 9  
US-10-282-122A-20027  
; Sequence 20027, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20027  
; LENGTH: 618  
; TYPE: DNA

```
; ORGANISM: Enterobacter cloacae
US-10-282-122A-20027

Query Match      29.8%; Score 357.4; DB 12; Length 618;
Best Local Similarity 75.9%; Pred. No. 1.6e-113;
Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY      557 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATTTTAAACGCTGTGCGCA 616
Db      1 ATGACCTTGAATGGTGGTTTGCCTACCTGCTGACATCGATTTTAAACGCTGTGCGCG 60

QY      617 GGCTCTGTCGAATCAACACTATGACCACTCGTCAACCGGTTA--TCCGGCCGGTG 674
Db      61 GGCTCGGGGCTATTAAACACCATGACCACTCCATCAATCACGGCTACCGGGGTGCGGG 120

QY      675 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTCATATGCTGCTGTTGGCGT--GGG 733
Db      121 GCATCGATTGCCGTTTGAGACCGGCTGGGTATTCATATGCTGCTGGGATCGGT 180

QY      734 GTTGGGACGCTATTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCGAGCGCG 793
Db      181 CTGGGACCCCTGTTCTCCCGCTCCGCTGGCTTGGGTGCTGAAATGGGCGGGGCA 240

QY      794 GCTTACTGATTGCTGGGGAATCCAGCAGTGGCGCGCTGCTGCTGCAATTGACCTTAAA 853
Db      241 GCGTATCTGATTGCTGCGCATCCAGCAGTGGCGCGGCTCCATCAACCTGAAT 300

QY      854 TCGTGGCCTCTACTCAATCGGTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTC 913
Db      301 ACGCTGGCTCTGACGCAAAACCGCGGCCATCTGTTTAAAGCTGCGGTATTCGTCAACCTG 360

QY      914 ACCAATCCCAAAAGATGTTGTTTCTGGCGGCGTATTTCCGCAATTCAATCATGCCGAA 973
Db      361 ACCAACCAGGAGAGATCGTGTTCCTCGCCGCGTGTTCGCCGAGTTTATCGTTCGGCAT 420

QY      974 CAGCCGCAACTGATGAGTATATCGTCTCGGCTCGGCTCACCACATATTGTTGCGATATT 1033
Db      421 CAGCCTCAGGTGATGAGTACGTTGCTGCTGGCGCGCAGCACTATTATCGTCGATATC 480

QY      1034 GTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG 1093
Db      481 GTGATGATTGGTTACCGGACGCTGGCGCAGCGAATTGCGGCGTGATTAAGGGCCCTAAG 540

QY      1094 CAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGTTA 1153
Db      541 CAGATGAAGGCCCTGAATAAAGCTTTGGCTCGCTGTTTATGCTGTTGCGCGCTGCTT 600

QY      1154 GCATCGCGGAGGCATGC 1170
Db      601 GCGTCAGCGCGTCAAGC 617

RESULT 10
US-10-282-122A-36679
; Sequence 36679; Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36679
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-36679

Query Match      29.0%; Score 348.2; DB 12; Length 623;
Best Local Similarity 76.6%; Pred. No. 2.6e-110;
Matches 479; Conservative 0; Mismatches 138; Indels 8; Gaps 4;

QY      557 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAATTTAAACGCTGTGCGCA 616
Db      1 ATGACCTTGAATGGTGGTTTGCCTATCTACTGACCTCAACTCTGCTGAGTCTTTCTCCG 60

QY      617 GGCTCTGTCGAATCAACACTATGACCACTCGCTCGCTCAACCAACCGTTATCCGGCG--GTG 674
Db      61 GGTTCAGGCGCCATCAATACCATGACGACGCTCTATCAACCATGGATATCGTGGCGCAGCG 120

QY      675 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATCGCATATATTGTCGCTGGTGGCGTGGGG 734
Db      121 GCTTCTATCGCCGACTCCAGACCGGACTGGGATACATATCGTACTGGTGGCGCTCGGA 180

QY      735 TTGGG--ACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCGAGCGCG 793
Db      181 CTGGGTACGCTCTTTTCGGGCTCGCTCCTCGCTTTTGAAATTTCTGAAATGGGCTGGCGCG 240

QY      794 GCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATATTGACCTTAA 853
Db      241 GCTTATCTTATCTGGCTGGGTATCCAGCAATGGCG--CGCAGCGCTATCGATCTGCAT 297

QY      854 TCGCTGGCCTCTACTCAATCGGTCGACATTTGTTCCAGCGCGCAGTTTTTTGTGTAATCTC 913
Db      298 ACTCTGCCAGACGCAATCGCGGGTGGGTGTTCAACCGGGCGATATTTGTCAATCTA 357

QY      914 ACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGGCAATTCATCATGCGGCAA 973
Db      358 ACCAATCCCAAAAGTATTGTCTTTTCTGGCCGCTGTTTCGGCAATTCATCATGCGGCGAG 417

QY      974 CAGCCGCAACTGATGCACTATATCGTCTCGGCGTCACCACTATTGTTGGTGCATATTATT 1033
Db      418 CAACCGCAACTGGCGCAGTACCTCATCTCGGCGTCACCACTATTGTTGGTGGATATGATT 477

QY      1034 GTGATGATCGGTTACGCCACCCCTTGTCTCAACGG--ATTGCTCTATGGATTAAAGGACCAA 1091
Db      478 GTGATGACCGGTTACGCCACACTGGCGCTAGCGCACTTGGCGCTGGGATTAAAGGACCAA 537

QY      1092 AGCAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGT 1151
Db      538 AGCAGATGAAGGCGCTGAATAAGCGTTTGGTTCGTTGTTTATGCTGGTAGGCGCGCTCC 597
```







```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23738
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23738

Query Match      23.0%; Score 276.4; DB 12; Length 990;
Best Local Similarity 76.2%; Pred. No. 4.3e-85;
Matches 340; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 446 AGAAAAAGCGTGGCAATCGCTCCGCCATAGAAATGGCAAGGATAAACCGTTTTCGCCAG 387

QY 61 GGACCGGGCTGAACCTCTGCTGCCAGATGCGGCCAGATGCGGCCAGATCAATCAATTAAG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 386 TGTCCAGGAACCACTGCTGCTGCCAGAGCGCGGCCAGGTCAATCAATAGTCTGCTGAAG 327

QY 121 CGATTAAATGCGCCGAGATGCGGATCGGCTAACAGCGACCGGAACGTCCTCCCGCGGA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 TTAACCAAGTGTCCCGGTGGGTATCTGAGAGCAGCGCGGCCCGAGTCCCTGCGCGGA 267

QY 181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTCAATAGGCCAGTTCCGCATAT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 TGGTCAATGATCATCACGTCAAAGCCGAGTGAAAGAGATCGTAGGCCACCTTCAGCATAC 207

QY 241 TTTACGTAGCTCTCAATACGCCCCGGGCGAGATGACTACCAACCGGTCAATGGTGTGTGG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 TTCACGTAGCTTCAATGCGCCCCGGGCAATCAACACAGACGATCGTTATGTTCGGCG 147

QY 301 CGAAACGGACAAAGCGCACCGGAATGTCATCCACACCAAGTAACTCTGCTTCATCAGCG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 CAAAGCGGACAAAGCGTACCGGTACATCGTCGACGCGCTTTAAATTCAAGTCTTCGCGC 87

QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGGTAAAGCAGCAAAACGCGTTTCTCTTGT 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86 TGTGCGCAGAAATCCGTCAGCGGCCCATGCTAAAGSCCGCAACGCGTTTCTCTGTGT 27

QY 421 TCCAGTCTTTTGTGCTGCTGAAACAT 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 TCCAGTCTTTTGTGCTGCTGAAACAT 1
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 15

```
US-10-282-122A-41840
; Sequence 41840, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41840
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-282-122A-41840

Query Match      22.9%; Score 275; DB 12; Length 621;
Best Local Similarity 67.3%; Pred. No. 9.9e-85;
Matches 418; Conservative 0; Mismatches 200; Indels 3; Gaps 2;

QY 557 ATGACCTTAGAATGGTGGTTTGCCCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGACCTTAGACTGGTGGTTTAAACCTATCTGCTGACAAACATTGATCCTCAGTCTTTCTCCC 60

QY 617 GGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACCACGGTTATC--CGGCCGGTG 674
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GGCTCGGGGGCCATTAAACAAATGAGCACTGCCATCAGCCATGGTACCCGTGGGTTGTT 120

QY 675 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGTGCTGGTTGGCGT-GGG 733
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GCTTCCATTGGCGGTTTACAACCTGGTCTGCGGTTGCAATATCGTCTTGGTTGGCGTGGT 180

QY 734 GTTGGGACGCTATTTCCTCGCTCAGTGATTCGCTTTTGAAGTGTGGAAGTGGGCAGCGCG 793
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CTGCTGCCCTGGTCTCTCAATCCTTACTGGCATTGGAATATTAAAGTGGTCTCGGTGG 240

QY 794 GCTTACTTGTATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCAATTGACCTTAAA 853
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GCTTACTTAATTGGCTTGGTATTCAACAATGGCGTGCAGSCCGGTGCGTTGACCTGCAT 300

QY 854 TCCTGGCCTCTACTCAATCGCGTTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 913
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GCACTGGCAACACAGTATGCCAGGCGTAAGCTGTTTAAACGTGCGGTGTTGTTAACTG 360

QY 914 ACCAATCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTTCATCATGCCGCA 973
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ACCAATCCAAAAGCATTGTTTCTGGCCGCGTTATTTCACAAATTTGTACTGCCACAG 420

QY 974 CAGCCGCAACTGATGAGTATATCGTGTGCGCGCTCACCACATATTGTGTGATATTATT 1033
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CAACCGCAGGTGGCACAGTATTGATTTTGGGCGAGCACCGAGTGTGATGTCGATATTATC 480
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 28, 2004, 14:07:39 ; Search time 3810.91 Seconds  
(without alignments)  
9403.168 Million cell updates/sec

Title: US-09-847-392-1  
Perfect score: 1200  
Sequence: 1 agaaataatgtggagatcgc.....atgtcggatgcggcgtaaac 1200

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vr1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	634.6	52.9	769	28	AQ858095 nbe0011L
C 2	371	30.9	473	28	AQ859250
C 3	339	28.2	687	28	AY080100
4	218.2	18.2	815	29	CNS011B AL149432 Anopheles

5	120.6	10.0	954	28	AF114214
6	85.4	7.1	1095	28	BZ548406
C 7	55.8	4.7	1199	28	BZ566001
C 8	53	4.4	231	14	CB042356
9	44.8	3.7	667	28	AQ991154
10	41.4	3.4	922	29	CNS014YP
C 11	41	3.4	849	13	EX462111
12	40.6	3.4	1039	29	CNS016GW
C 13	39.8	3.3	475	28	BH209707
14	39.8	3.3	990	29	AY401573
15	39.4	3.3	712	13	EX416727
16	39.4	3.3	885	13	EX425603
17	39.2	3.3	1101	29	CNS002LD
18	39	3.2	588	14	CB281433
C 19	38.8	3.2	1146	13	EX446722
20	38.8	3.2	685	13	BY724027
C 21	38.8	3.2	802	29	EX167575
22	38.8	3.2	1101	29	CNS00BSO
C 23	38.6	3.2	642	9	AL628057
24	38	3.2	510	12	BG664436
C 25	38	3.2	1201	13	EX376097
C 26	37.8	3.1	1101	29	CNS017QV
27	37.8	3.1	1201	9	AL567139
C 28	37.6	3.1	798	10	BE659545
C 29	37.4	3.1	586	14	CD651858
30	37.2	3.1	1159	29	CNS015XR
C 31	37.2	3.1	1201	13	EX381961
C 32	37.2	3.1	1201	13	EX394709
33	37	3.1	264	10	AW347581
34	37	3.1	308	10	AW266920
35	37	3.1	437	12	BG688576
36	37	3.1	450	10	AW482031
37	37	3.1	495	10	BF606721
38	37	3.1	516	10	BE663253
39	37	3.1	522	12	BM087647
40	37	3.1	539	12	BM253190
41	37	3.1	573	10	BE665319
42	37	3.1	573	12	BI773985
43	37	3.1	585	12	BI774021
44	37	3.1	599	14	CB466945
45	37	3.1	603	9	AV596505

ALIGNMENTS

RESULT 1  
AQ858095  
LOCUS  
DEFINITION  
769 bp DNA linear GSS 03-NOV-1999  
nbe0011L14f CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica cultivar-group) genomic clone nbe0011L14f, genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
AQ858095  
AQ858095.1 GI:6208552  
GSS.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 769)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGGG  
Class: BAC ends







/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="18P02"  
/clone\_lib="NotreDame1"  
/note="end : T7"

ORIGIN

Query Match 18.2%; Score 218.2; DB 29; Length 815;  
Best Local Similarity 77.0%; Pred. No. 5e-55;  
Matches 298; Conservative 4; Mismatches 82; Indels 3; Gaps 3;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db |||||  
423 AAAAACAGCGTTGAATCGCACCGCCCATGTAGTGGCCAGAAATATAGCGTTCCGCCAG 482  
QY 61 GGACCGGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120  
Db |||||  
483 GGGCCTGGCTCGACCTCTCTGCTGCCAGACGGCGGAGATCGTCGACGTAATCGCTGAAG 542  
QY 121 CGATTACATGCCCCGAGATCGGGATCGGCTAACAGGGACCGGACGTCCTGCCCGCG- 179  
Db |||||  
543 TTATCCACGTGGCCGGATGGGTATCCGATAACAGGGCCCTGAARCCCTGCCCGCGS 602  
QY 180 ATGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTTCATAGGCCAGTTCCGCATA 239  
Db |||||  
603 GTGATCGATAATCAGCACATCGAAGCCAGATGGAACAGGTTCATAGGCCAGTTCCGCCGTA 662  
QY 240 TTTTACGTAGTCTCAATAGCCCCCGGGCAGATGACTACCAACCGTTCATGGTGTGTC 299  
Db |||||  
663 TTTKATGTAGTCTCAATAGCTCCGGGCGAGACCACATGACCCCGTCAATTTTCTTATC 722  
QY 300 GCGAAACGGCAAAAGCGCACCGGAATGTCATCCACACAGTAACCTGCTTCATCACG 359  
Db |||||  
723 GTG-AAAMGCAAAAGCGTACTGGAATATCGCGGACGCCCGTAACTCATCTTCTCAGG 781  
QY 360 CTGACGCCAGAAATCAGTCAGCGGTCC 386  
Db |||||  
782 CTG-CGCTAGAAATCGGTCAGCGSCC 807

RESULT 5  
AF114214  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AF114214 954 bp DNA linear GSS 29-AUG-2000  
AF114214 Salmonella typhimurium LT2, Lambda DASH II Salmonella  
typhimurium genomic clone 415-T3, genomic survey sequence.  
AF114214  
AF114214.1 GI:5281261  
GSS.  
Salmonella typhimurium  
Salmonella typhimurium  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
1 (bases 1 to 954)  
Wong,R.M.Y. and McClelland,M.  
End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,  
Li-Cor  
Unpublished (1999)  
Contact: McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3099 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@lifsci.sdsu.edu  
Class: shotgun.  
Location/Qualifiers  
1. .954  
/organism="Salmonella typhimurium"  
/mol\_type="genomic DNA"  
/strain="LT2"  
/db\_xref="taxon:602"  
/clone="415-T3"  
/clone\_lib="Salmonella typhimurium LT2, Lambda DASH II"

FEATURES  
source

Query Match 7.1%; Score 85.4; DB 28; Length 1095;  
Best Local Similarity 54.1%; Pred. No. 3.2e-14;  
Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;

QY 774 TGTGAAGTGGCAGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGC--CGGC 831  
Db |||||  
41 TGATCAAGTGGTTCGGCGTGGCCTACCTGTGTACCTGTGTCGGCGGTGGCAGGCGC 100  
QY 832 CGTGGTGCATTTGACCTTAAATCGTGGCTCTACTCAATCG-CGTGACATTTCTCC 890  
|||

/note="Vector: Lambda DASH II; sequenced using Li-Cor  
sequencer"

ORIGIN

Query Match 10.0%; Score 120.6; DB 28; Length 954;  
Best Local Similarity 77.0%; Pred. No. 4.7e-25;  
Matches 147; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60  
Db |||||  
763 AGGAACAGGTGGCTATCGCGCCCCCATGTAGTCCGCCAGGATATACGCTTACGCCAC 822  
QY 61 GGACCGGGCTGAACCTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120  
Db |||||  
823 GGCCCCGGCTCAATTTCTGTGTCAGAACGCCGCTAAGTCTCAACATAATCGTTAAAAA 882  
QY 121 CGATTACATGCCGAGATCGGATCGGCTAAACAGGCGGACCGGAAACGTCCTGCCCGGA 180  
Db |||||  
883 TGATCGACATGACCCCGATGGGGTCCGATAACATCCGCCGGAGCGCCCTTGACCACGG 942  
QY 181 TGGTCGATGAT 191  
Db |||||  
943 TGGTCGATGAT 953

RESULT 6  
BZ548406  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BZ548406 1095 bp DNA linear GSS 17-DEC-2002  
pacsl-60\_1079.s1 pacsl-60 Pseudomonas aeruginosa genomic clone  
pacsl-60\_1079, genomic survey sequence.  
BZ548406  
BZ548406.1 GI:27151987  
GSS.  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 1095)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers  
1. .1095  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pacsl-60\_1079"  
/clone\_lib="pacsl-60"  
/note="clinical isolate 1-60 Whole genomic shotgun  
library."

FEATURES  
source

Query Match 7.1%; Score 85.4; DB 28; Length 1095;  
Best Local Similarity 54.1%; Pred. No. 3.2e-14;  
Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;

QY 774 TGTGAAGTGGCAGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGC--CGGC 831  
Db |||||  
41 TGATCAAGTGGTTCGGCGTGGCCTACCTGTGTACCTGTGTCGGCGGTGGCAGGCGC 100  
QY 832 CGTGGTGCATTTGACCTTAAATCGTGGCTCTACTCAATCG-CGTGACATTTCTCC 890  
|||

Db 101 CGCCACAGGCCTTGAGCACCGGATGGCGAACGGCCTCTCTGGGGCGACCGTTGACCCCTGGTGC 160

QY 891 AGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAGATATTGTGTTCTGGCGCGCTAT 950

Db 161 TGCCTGTTTCTCTGTTCAAGCCAGCAATCCCAAGCGGTGATCTTCATGCTCGCGTGC 220

QY 951 TTCCGCAATTCATCATGCGCGCAACAGCCGCAACTGATGTCAGTATATCGTCTCGCGCTCA 1010

Db 221 TGCCGCAGTTTCATCGACCCGCAACAGCGCTGCTGGCGCAATACCTGATCATGGCGGCA 280

QY 1011 CCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGATTG 1070

Db 281 CCATGATCGTCTGTCGACCTGATCGTCATGGCCGGCTACACCGGCTGGCTGGCGCGTGC 340

QY 1071 CTCATGGAATTAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTTCGGCTCGTTGT 1130

Db 341 TACGCGTACTGCGTTTCGCGCGCCAGCAGAAGCTGGTGAACCGTACCTTCGCGAGCCTGT 400

QY 1131 TTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGCATGCG 1171

Db 401 TCGTCGTGTCGCGCGGGGTGCTGGCGACGGTACGCCGAGCG 441

RESULT 7

BZ566001/c

LOCUS

DEFINITION

BZ566001

pacS2-164\_5858.xl pacS2-164 Pseudomonas aeruginosa genomic clone

pacS2-164\_5858, genomic survey sequence.

ACCESSION

BZ566001

VERSION

BZ566001.1 GI:27194705

SOURCE

GSS.

ORGANISM

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE

1 (bases 1 to 1199)

AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE

Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library

JOURNAL

J. Bacteriol. (2002) In press

COMMENT

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

source

1. .1199

Location/Qualifiers

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacS2-164\_5858"

/clone\_lib="pacS2-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 4.7%; Score 55.8; DB 28; Length 1199;

Best Local Similarity 50.1%; Pred. No. 4e-05;

Matches 195; Conservative 0; Mismatches 187; Indels 7; Gaps 2;

QY 684 GCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTGGCGTGGGTTGGGACGC 743

Db 649 GCTGGCCCTCAGATCGGGCTGCCCTGCACATACCCATTGTCTCTGCCGGGTGGTACG 590

QY 744 TATTTTCC--CGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGCAGGCGCGCTTACTT 801

Db 589 TTACTGGCGACCGTCGGCACTGGCTTACAGCATGATCAAGTGGTTCGGCGGTGCCTACAT 530

QY 802 GATTTGGCTGGGAATCCAGAGTGGC-----GCGCCGCTGGTGTCAATTGACCTTAAATCG 856

Db 529 GGTGTACCTGCTGGTGGCCAGTGGCAGGCGCCGCCACAGACCTTGAGCACCGATGGCG 470

QY 857 CTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTTGTGAATCTCACC 916

Db 469 ACCGGCCCTCTGGGTGACACCGTTGACCCCTGGTGTGCTGGTGGTCACTCGGTGACCGACAC 410

QY 917 AATCCCAAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCAATTTCATCATGCCGCAACAG 976

Db 409 AATCACAAGACGGTGTCTGTCATGCTCGCGGTGCGTCCCCAGTTTCAGGACCGGCATCAT 350

QY 977 CCGCAACTGATGTCAGTATATCGTGTCTCGCGGTCAACCATATTGTGGTCGATATTATTGTG 1036

Db 349 CCGCTGCTGGCGCAAGACCTGATCATGGCGGCCCCCATGATCGTCTGTCGTGTCGTCGTC 290

QY 1037 ATGATCGGTTAGCCACCCCTTGCTCAACG 1065

Db 289 ATGTCGAGTACACCGGGTGGCTGCGCG 261

RESULT 8

CB042356/c

LOCUS

DEFINITION

CB042356

4007198 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1P12

Unknown, mRNA sequence.

ACCESSION

CB042356

VERSION

CB042356.1 GI:27761601

KEYWORDS

EST.

SOURCE

Bos indicus (zebu)

ORGANISM

Bos indicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 231)

AUTHORS

da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K., Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,B.E., Machado,M.A. and Coutinho,L.L.

TITLE

Construction and Characterization of cDNA Libraries Generated from Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos indicus) Cattle

JOURNAL

Unpublished (2002)

COMMENT

Contact: Adilson F. da Mota

Gene Evaluation and Mapping Laboratory

USDA, ARS, Animal and Natural Resources Institute

Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA

Tel: 3015048456

Fax: 3015048414

Email: amota@cnpgl.embrapa.br

Single pass sequencing. Bases called and trimmed with phred 0.000925 using options -trim\_alt '' -trim\_fasta. Vector identified by cross\_match using options -minmatch 12 -minscore 18

PCR Primers

FORWARD: GTTTTCCCAGTCACGACGTTG

BACKWARD: TGAGCGGATACAAATTTCCACAG

Plate: 1 row: P column: 12

Seq primer: GTTTTCCCAGTCACGACGTTG

High quality sequence stop: 231.

FEATURES

Location/Qualifiers

1. .231

/organism="Bos indicus"

/mol\_type="mRNA"

/strain="Brazilian Dairy Gir"

/db\_xref="taxon:9915"

/clone="316BOV\_1P12"

/sex="female"

/tissue\_type="mammary"

/cell\_type="epithelium"

/dev\_stage="involutd"

/lab\_host="X-12"

/clone\_lib="BARC-EMBRAPA 316BOV"

/note="Organ: mammary; Vector: pUC19; Site:1: SmaI; Site\_2: SmaI; This mammary-derived cDNA library was



Db 748 WYHTWYCCYGYCCGAAAAWBKTNRTGABSSARRCGKCVBSKYSVNVYRDVAAAMCSC 807  
QY 543 ACCGGAGTTCATCATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATGCATCATTT 602  
Db 808 SAMCCCCMRKCKSCYCNWARDAGADMYSBCKTABDKTKTRWKGSGTAMCKCCCCMYK 867  
QY 603 T 603  
Db 868 T 868  
RESULT 11  
EX462111/c  
LOCUS  
DEFINITION  
BX462111 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
clone CS0DG001YP07 5-PRIME, mRNA sequence.  
ACCESSION  
EX462111  
VERSION  
EX462111.1 GI:31029391  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 849)  
AUTHORS  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE  
Full-length cDNA libraries and normalization  
JOURNAL  
Unpublished (2001)  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9966.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DG001CH04QPI&cluster=9966.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DG001CH04QPI.  
FEATURES  
Location/Qualifiers  
1..849  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DG001YP07"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
ORIGIN  
Query Match 3.4%; Score 41; DB 13; Length 849;  
Best Local Similarity 21.3%; Pred. No. 1.2;  
Matches 53; Conservative 86; Mismatches 110; Indels 0; Gaps 0;  
QY 661 TTATCCGGCGGTGGCGTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGTGC 720  
Db 500 TTGTGCTGCTTTTGCTGCTTTGCTTAGCNTTTTGGCGCTATTGCGAGCGCGCATTTGC 441  
QY 721 TGGTTGGCGTGGGTTGGGACGCTATTTCCCGCTCAGTGATTCGGTTTGAAGTGTGAA 780  
Db 440 TTGCKTGCKKGGCKGCKKCKKSGCKTGCGCKKGGCTSTTSTSKSKSKSGSKGS 381  
QY 781 GTGGCAGCGCGGCTTACTTGATTGGCTGGGAATCCAGACGTGGCGCGCTGGTGC 840  
Db 380 KKSSTSASGSKTGSSKKAKKSSKKKSAKSGSKGSKGSKGSKGSKGSKGSNS 321  
QY 841 AATTGACCTTAAATCGTGGCCTCTACTCAATCGGCTGACATTTGTTCCAGCGGCAGT 900

Db 320 TGSNSCNSKSKSAKTATSGSGSKKKGSKTASTTSKKKSKSKKSKSTSKKSKSKSKKA 261  
QY 901 TTTTGTGAA 909  
Db 260 CKGSKKKA 252  
RESULT 12  
CNS016GW  
LOCUS  
DEFINITION  
CNS016GW 1039 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN16011 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
AL106730  
VERSION  
AL106730.1 GI:5623541  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1039)  
REFERENCE  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.  
FEATURES  
Location/Qualifiers  
1..1039  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN16011"  
/clone\_lib="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : T7"  
ORIGIN  
Query Match 3.4%; Score 40.6; DB 29; Length 1039;  
Best Local Similarity 30.6%; Pred. No. 1.7;  
Matches 57; Conservative 54; Mismatches 75; Indels 0; Gaps 0;  
QY 628 AATCAACACTATGACCACCTCGCTCAACACGGTTATCCGGCGGTGCGTCTATTGCTG 687  
Db 829 WAAAAAAMAAMWTHMARRGTGMAWMWTTGGGGGTTTGGKGGTTCMVGSMTM 888  
QY 688 GGCTTCAGACCGGACTGGCGATTCAATGTGCTGGTGGCGTGGGTTGGACGCTATT 747  
Db 889 KGTTTTBGGVGGTTTGTGTVKTKTHDGTGTCCTTBYCCKSGSGTTGKYCYCCCYT 948  
QY 748 TTCCCCTCAGTGATTGCGTTTGAAGTGTGAAGTGGCAGGCGCGCTTACTTGATTG 807  
Db 949 BYCCGGKTTTCTTCHTGGKTTKGGTCKTGGGKGTTCNKKKGCVHTTTKKKKKKKK 1008  
QY 808 GCTGGG 813  
Db 1009 KTKKGG 1014  
RESULT 13  
BH209707/c  
LOCUS  
DEFINITION  
BH209707 475 bp DNA linear GSS 15-NOV-2002  
Sml-41P23.TF Sml Schistosoma mansoni genomic clone Sml-41P23,  
genomic survey sequence.

ACCESSION BH209707  
VERSION BH209707.1 GI:16388592  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni  
ORGANISM Schistosoma mansoni  
REFERENCE Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
AUTHORS Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.  
1 (bases 1 to 475)  
Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and El-Sayed,N.M.  
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml  
JOURNAL BAC library for gene discovery and map construction  
COMMENT Unpublished (2001)  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml  
BAC library. For clone availability, please contact Dr. Najib  
El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip Loverde at  
State University of New York, Buffalo, New York, USA  
(loverde@buffalo.edu)  
Seq primer: M13 For  
Class: BAC ends.

FEATURES  
source  
1..475  
/organism="Schistosoma mansoni"  
/mol\_type="genomic DNA"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sml-41P23"  
/clone\_lib="Sml"  
/note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed  
in the laboratory of Dr. Denis Le Paslier at the Fondation  
Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma  
mansoni agarose embedded DNA was partially digested with  
Hin dIII. High molecular weight fragments were ligated in  
pBelobAC11 digested with Hin dIII. The average insert size  
is 100 kb. Total clone coverage: approx. 7.95 X the  
haploid genome. Further information can be found in Le  
Paslier et al. (2000) Construction and characterization of  
a Schistosoma mansoni bacterial artificial chromosome  
library. Genomics 65: 87-94."

ORIGIN  
Query Match 3.3%; Score 39.8; DB 28; Length 475;  
Best Local Similarity 46.8%; Pred. No. 1.9;  
Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 933 TGTTCCTGGCGGCTATTTCGCAATTCATCATGCCGCAACAGCGCAACTGATGCAGT 992  
Db 352 TGATCCTGCTGTGCTGATGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGC 293  
QY 993 ATATCGTCTCGGCGTCACCACTATTGCGTCGATATTATTGTGATGATCGGTACGCCA 1052  
Db 292 TGCTGCTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGA 233  
QY 1053 CCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATA 1112  
Db 232 TGATGATGATGATGATGATGATGAAGATGATGTGGATGAAGTTGAAGAAGGGATGATA 173  
QY 1113 AGATTTTCGGCTGTTGTTTATGCTGTGGGAGCGCTGTATTAGCATCGCGGAGGATCGGT 1172  
Db 172 AGATTTTGGACATGATGTGTTTGTGACGATGACGATGATGATGATGATGATGAAGATAA 113  
QY 1173 GAAAAATAATGTCGGATCGGGGTAAA 1199  
Db 112 GGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 86

RESULT 14  
AY401573  
LOCUS  
DEFINITION Mus musculus GPR7 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY401573  
VERSION AY401573.1 GI:39757562  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 990)  
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 990)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
1..990  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/locus\_tag="GPR7"  
/locus\_tag="HCM0935"

ORIGIN  
Query Match 3.3%; Score 39.8; DB 29; Length 990;  
Best Local Similarity 52.1%; Pred. No. 2.9;  
Matches 112; Conservative 0; Mismatches 102; Indels 1; Gaps 1;  
QY 580 CTACCTGCTGACATCGATCAATTTAAACGCTGTGCGCAGGCTCTGGTGCATCAACACTAT 639  
Db 337 CTCATTGTAGCCGTCGACCACTACAACTTCTCTAGCCTCTACTTC-CTGCCGTCAT 395  
QY 640 GACCACCTCGCTCAACACACCGTTATCCGCCGGTGGCGTCTATTGTGGGCTTCAGACCG 699  
Db 396 GAGCGCCGACCATACCTGCTGTTCTGCGCACAGACAGTCGCCCGGGTGTCCGGCG 455  
QY 700 GACTGGCGATTCAATATTGTGCTGGTTGGCTGGGTTGGGACGCTATTTCCCGCTCAGT 759  
Db 456 CACTTACGGTGCAGCGCGTGTCTCAGTCTGCGGTGTGGCGCTGTGGTACGCTGGTCTGT 515  
QY 760 GATTCGCTTTGAAGTGTGAAGTGGGACGCGCGG 794  
Db 516 GCTGCCCTTTGCGGTATTCGCTCGGCTGGACGAGG 550

RESULT 15  
BX416727  
LOCUS  
DEFINITION BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
CS0DA011Y114 5-PRIME, mRNA sequence.  
ACCESSION BX416727  
VERSION BX416727.1 GI:30765629  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 28, 2004, 14:04:39 ; Search time 2572.83 Seconds  
(without alignments)  
10360.543 Million cell updates/sec  
Title: US-09-847-392-1\_COPY\_557\_1171  
Perfect score: 615  
Sequence: 1 atgaccttagaatggtggtt.....tagcatcgcgaggcatgcg 615

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rtd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	615	100.0	1200	6	AR172781	AR172781 Sequence
2	615	100.0	1200	6	E41573	E41573 DNA encodin
3	615	100.0	1231	6	E48927	E48927 Novel gene
4	615	100.0	1231	6	AX030175	AX030175 Sequence
C 5	604	98.2	91414	1	ECOUW85	M87049 E. coli gen
C 6	595	96.7	11509	1	AE000458	AE000458 Escherich
C 7	577.6	93.9	10253	1	AE015396	AE015396 Shigella
C 8	571.2	92.9	290976	1	AE016990	AE016990 Shigella
C 9	571.2	92.9	318703	1	AP002567	AP002567 Escherich
C 10	566.4	92.1	301566	1	AE016769	AE016769 Escherich
C 11	553.8	90.0	10592	1	AE005613	AE005613 Escherich
C 12	467	75.9	212936	2	AC020970	AC020970 Mus muscu
C 13	449.8	73.1	256373	2	AC020870	AC020870 Mus muscu
14	414	67.3	417	6	AR352333	AR352333 Sequence
15	379.2	61.7	258050	1	AL627278	AL627278 Salmonell
16	379.2	61.7	299991	1	AE016845	AE016845 Salmonell
17	375	61.0	645	6	AR384166	AR384166 Sequence
C 18	374.4	60.9	21692	1	AE008884	AE008884 Salmonell
C 19	374.4	60.9	96086	1	STYSTMD1	AF233324 Salmonell
C 20	358.8	58.3	265383	2	AC020874	AC020874 Mus muscu
C 21	274.8	44.7	10957	1	AE013640	AE013640 Yersinia
22	274.8	44.7	199050	1	AJ414159	AJ414159 Yersinia
23	241.4	39.3	268294	2	AC020885	AC020885 Mus muscu
C 24	207.4	33.7	263	6	AR352166	AR352166 Sequence
25	182.8	29.7	62274	2	AC020833	AC020833 Mus muscu
26	170.4	27.7	206758	2	AC016129	AC016129 Drosophil
27	162.6	26.4	299938	1	AE016800	AE016800 Vibrio vu
28	161	26.2	245560	1	AP005330	AP005330 Vibrio pa
C 29	151.4	24.6	288108	1	AP005083	AP005083 Vibrio pa
30	149.8	24.4	11305	1	AE004104	AE004104 Vibrio ch
31	123.6	20.1	3131	1	AB105408	AB105408 Photobact
C 32	112.4	18.3	10475	1	AE004937	AE004937 Pseudomon
33	109.4	17.8	110000	2	AC074221_2	Continuation (3 of
34	104	16.9	313518	1	AE016856	AE016856 Pseudomon
35	86.4	14.0	303226	1	AE016774	AE016774 Pseudomon
36	81.6	13.3	255260	1	BX640436	BX640436 Bordetell
37	80	13.0	123385	1	BX640452	BX640452 Bordetell
38	80	13.0	346359	1	BX640411	BX640411 Bordetell
39	75.4	12.3	663	6	AR318227	AR318227 Sequence
40	74.8	12.2	345783	1	AP003001	AP003001 Mesorhizo
C 41	72.8	11.8	63847	2	AC101459	AC101459 Mus muscu
C 42	72.2	11.7	346897	1	AP002995	AP002995 Mesorhizo
C 43	70.6	11.5	13301	1	AE005614	AE005614 Escherich
44	70	11.4	349737	1	BX572597	BX572597 Rhodopseu
45	61.8	10.0	11501	1	AE009276	AE009276 Agrobacte

ALIGNMENTS

RESULT 1	AR172781	AR172781	1200 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	Sequence 1	from patent US 6303348.				
DEFINITION	Sequence 1	from patent US 6303348.				
ACCESSION	AR172781					
VERSION	AR172781.1	GI:17912272				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1200)					
AUTHORS	Livshits,V.Arkadiyevich., Zakataeva,N.Pavlovna.,					
	Aleoshin,V.Venyamiyovich., Balareova,A.Valentinovna. and					
	Tokhmakova,I.Lvovna.					
TITLE	DNA coding for protein which confers on bacterium escherichia coli					



RESULT 3  
E48927  
LOCUS E48927 1231 bp DNA linear PAT 31-JAN-2002  
DEFINITION Novel gene and process for producing amino acid.  
ACCESSION E48927  
VERSION E48927.1 GI:18627996  
KEYWORDS JP 2000189177-A/1.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 1231)  
AUTHORS Rivishittsu,V.A., Zakataeva,N.P., Aryoshin,V.V., Beraryo,A.V. and Tokumakova,I.R.  
TITLE Novel gene and process for producing amino acid  
JOURNAL Patent: JP 2000189177-A 1 11-JUL-2000;  
AJINOMOTO CO INC  
COMMENT OS Escherichia coli  
PN JP 2000189177-A/1  
PD 11-JUL-2000  
PF 15-DEC-1999 JP 1999356018  
PR 23-DEC-1998 RU 98123511  
PI VITARI ARUKAJEVICHI RIVISHITTU,NATARIYA PAVUROVUNA ZAKATAEVA.  
PI VLADIMIR VENIYAMINOVICHI ARYOSHIN,ARA VALENTINOVUNA BERARYO,  
PI IRINA RIVOVUNA TOKUMAKOVA  
PC C12N15/09,C07K14/245,C12N1/21,C12P13/06,C12P13/08,  
PC C12P13/08//  
PC (C12N15/09,C12R1:19),(C12P13/06,C12R1:19),(C12P13/08,C12R1:19), PC  
C12N15/00,  
PC (C12N15/00,C12R1:19)  
CC  
FH Key Location/Qualifiers  
FT CDS (557)..(1171).  
FEATURES  
source  
1..1231  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
ORIGIN  
Query Match 100.0%; Score 615; DB 6; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 1.9e-185;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACCTTAGAATGGTGGTGGCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 60  
DB 557 ATGACCTTAGAATGGTGGTGGCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 616  
QY 61 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCACTGATCCGGCGGTGGC 120  
DB 617 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCACTGATCCGGCGGTGGC 676  
QY 121 GTCATTGCTGGGCTTCAGACCGGACTGGCGGCTGATTCATTTGCTGGTGGCGGT 180  
DB 677 GTCATTGCTGGGCTTCAGACCGGACTGGCGGCTGATTCATTTGCTGGTGGCGGT 736  
QY 181 GGGACGCTATTTCCCGCTCAGTCGATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGT 240  
DB 737 GGGACGCTATTTCCCGCTCAGTCGATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGT 796  
QY 241 TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGAATGACCTTAATCG 300  
DB 797 TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGAATGACCTTAATCG 856  
QY 301 CTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACC 360  
DB 857 CTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACC 916  
QY 361 AATCCCAAAAGTATTTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCCGCAACAG 420

DB 917 AATCCCAAAAGTATTTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCCGCAACAG 976  
QY 421 CCGCAACTGATGCAGTATATCGTCTCGGCTCGGCTCACCACCTATTTGTCGATATTATTGTG 480  
DB 977 CCGCAACTGATGCAGTATATCGTCTCGGCTCGGCTCACCACCTATTTGTCGATATTATTGTG 1036  
QY 481 ATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAG 540  
DB 1037 ATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAG 1096  
QY 541 ATGAAGGCGCTGAATAAGATTTCGGCTCGTGTGTTTATGCTGGTGGGAGCGCTGTTAGCA 600  
DB 1097 ATGAAGGCGCTGAATAAGATTTCGGCTCGTGTGTTTATGCTGGTGGGAGCGCTGTTAGCA 1156  
QY 601 TCGCGGAGGCATGCG 615  
DB 1157 TCGCGGAGGCATGCG 1171  
RESULT 4  
AX030175  
LOCUS AX030175 1231 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent EP1013765.  
ACCESSION AX030175  
VERSION AX030175.1 GI:10190392  
KEYWORDS Escherichia coli  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1  
AUTHORS Belareva,A.V., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L. and Zakataeva,N.P.  
TITLE Gene and method for producing l-amino acids  
JOURNAL Patent: EP 1013765-A 1 28-JUN-2000;  
AJINOMOTO KK (JP)  
FEATURES  
source  
1..1231  
/organism="Escherichia coli"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:562"  
557..1174  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC09318.1"  
/db\_xref="GI:10190393"  
/translation="MTLEWFWFAYLLTSIIILTSPGSGAINTMTSLNHGYPAGGVYCW  
ASDRTGDSYCGAGRWGVTLFPSRVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLA  
STQSRHLFQRAVFVNLNPKSIVFLAALFPIMPQQLMQYIVLGVTTIVVDIIV  
MIGYATLIAQRIALWIKPKQMKALNKIFGSLFVLVGLLASARHA"  
ORIGIN  
Query Match 100.0%; Score 615; DB 6; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 1.9e-185;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACCTTAGAATGGTGGTGGCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 60  
DB 557 ATGACCTTAGAATGGTGGTGGCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 616  
QY 61 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCACTGATCCGGCGGTGGC 120  
DB 617 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCACTGATCCGGCGGTGGC 676  
QY 121 GTCATTGCTGGGCTTCAGACCGGACTGGCGGCTGATTCATTTGCTGGTGGCGGT 180  
DB 677 GTCATTGCTGGGCTTCAGACCGGACTGGCGGCTGATTCATTTGCTGGTGGCGGT 736  
QY 181 GGGACGCTATTTCCCGCTCAGTCGATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGT 240  
DB 737 GGGACGCTATTTCCCGCTCAGTCGATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGT 796

QY 241 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTAATGACCTTAATCG 300  
Db |||||  
797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTAATGACCTTAATCG 856  
QY 301 CTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 360  
Db |||||  
857 CTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 916  
QY 361 AATCCAAAAGTATTGTGTTTTCGGCGCGCTATTTCGCAATTCATCATGCCGCAACAG 420  
Db |||||  
917 AATCCAAAAGTATTGTGTTTTCGGCGCGCTATTTCGCAATTCATCATGCCGCAACAG 976  
QY 421 CCGCACTGATGCGATATATCGTGCTCGCGTCACCACTATTGTGTCGATATTATTGTG 480  
Db |||||  
977 CCGCACTGATGCGATATATCGTGCTCGCGTCACCACTATTGTGTCGATATTATTGTG 1036  
QY 481 ATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGATTAAGGACCAACAG 540  
Db |||||  
1037 ATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGATTAAGGACCAACAG 1096  
QY 541 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCA 600  
Db |||||  
1097 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCA 1156  
QY 601 TCGGCGAGGCATCGG 615  
Db |||||  
1157 TCGGCGAGGCATCGG 1171

RESULT 5  
ECOUW85/c  
LOCUS  
DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.  
ACCESSION M87049  
VERSION M87049.1 GI:836656  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 91414)  
Daniels,D.L., Plunkett,G. III, Burland,V. and Blattner,F.R.  
Analysis of the Escherichia coli genome: DNA sequence of the region  
from 84.5 to 86.5 minutes  
Science 257 (5071), 771-778 (1992)  
92358234  
PUBMED 1379743  
REFERENCE 2 (bases 1 to 91414)  
Plunkett,G. III, Burland,V., Daniels,D.L. and Blattner,F.R.  
Analysis of the Escherichia coli genome. III. DNA sequence of the  
region from 87.2 to 89.2 minutes  
Nucleic Acids Res. 21 (15), 3391-3398 (1993)  
93347969  
PUBMED 8346018  
REFERENCE 3 (bases 1 to 91414)  
Blattner,F.R., Burland,V., Plunkett,G. III, Sofia,H.J. and  
Daniels,D.L.  
Analysis of the Escherichia coli genome. IV. DNA sequence of the  
region from 89.2 to 92.8 minutes  
Nucleic Acids Res. 21 (23), 5408-5417 (1993)  
94089392  
PUBMED 8265357  
REFERENCE 4 (bases 1 to 91414)  
Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. III, Lazar,S. and  
Rouviere,P.E.  
A new family of peptidyl-prolyl isomerases  
Trends Biochem. Sci. 20 (1), 12-14 (1995)  
95184296  
PUBMED 7878731  
REFERENCE 5 (bases 1 to 91414)  
Daniels,D.L.  
Direct Submission

JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax:  
608-263-7459  
COMMENT On May 29, 1995 this sequence version replaced gi:148169.  
This sequence was determined as part of the E. coli Genome Project  
(Frederick R. Blattner, director) at the University of  
Wisconsin-Madison. The entire sequence was independently determined  
from E. coli MG1655. Overlaps and conflicts with other sequence  
determinations are annotated. Reference [1] describes the original  
sequence determination of a 91408 bp sequence. References [2], [3],  
and [4] describe subsequent corrections and/or updates to that  
sequence.

FEATURES  
source Location/Qualifiers  
1..91414  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K-12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:562"  
/map="bp 1 at 3974 kb; 84.7 minutes"  
/clone\_lib="lambda library of Daniels and Blattner"  
/note="This sequence comprises the following lambda  
clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177),  
DD856(EC19RM8.1), DD858(EC17RM9),  
DD861(EC17-221),DD864(EC27-149), DD865(EC21-76),  
DD867(EC27-860). M13mp19 or Janus vectors were used for  
subcloning."  
misc\_feature 1..5863  
/note="clone EC15-116"  
misc\_feature complement(1..4899)  
/note="corresponds to M37337 (1..4900)"  
misc\_feature 1..339  
/note="corresponds to V00349 (164..500)"  
misc\_feature 1..200  
/note="corresponds to K02846 (1..200)"  
10..86  
/gene="aspT"  
10..86  
/gene="aspT"  
/note="CGSC No. 989"  
95..170  
/gene="trpT"  
95..170  
/gene="trpT"  
/note="CGSC No. 66"  
177..205  
/note="terminator-like sequence"  
misc\_structure 253..396  
/note="predicted bend of 81 degrees"  
gene complement(266..862)  
/gene="f198"  
CDS complement(266..862)  
/gene="f198"  
/function="unknown"  
/cds\_start=1  
/transl\_table=11  
/protein\_id="AA067566.1"  
/db\_xref="GI:148170"  
/translation="MAHTRHNEFSIGASLWECMLNQWLGRLYQNQDAHTGLQFEA  
RIAQRQSLVKQLHERQLDLITTEAPKMDFFSSQLGYFTLYTSAPSKLGDLYL  
RLWGPDPFQOHEAGLIGADEVPILITSSAELAQCIAMNLGCTWLPVSWARKKGLHT  
VVDSTLSRPLYAIWLQNSDKNALIRDLKINVLDEV"  
misc\_structure 586..724  
/note="predicted bend of 77 degrees"  
gene complement(703..1104)  
/gene="f133"  
CDS complement(703..1104)  
/gene="f133"  
/note="CGSC No. 18010; possible psr gene"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA067567.1"



```
/db_xref="GI:148171"
/translation="MDTELLIKTFLEVSRTFRHFGRAESLYLTQSAVSFRIRQLENQLG
VNLFTRRNNIRLTAAAGEKLLPYAETIMSTWQAARKRWIRPHDITSLFVPAAPRCGNV
CLISGWDACIKIRMPQIAQYSSKRELPLNGSLW"
complement(1136..1165)
/note="promoter matrix score of 63"
1151..1179
/note="promoter matrix score of 57"
1223..1236
/gene="o137"
1223..1236
/gene="o137"
1223..1236
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="AAA67568.1"
/db_xref="GI:148172"
/translation="MAESFTTNNRYFDNKHYPGRFSRHGDFTIKEAQLLERHGVAFNE
LDLGKREPTEEEKLFVAVCRGEREPVEAERVWSKYMTRIKRPFHTLSAVNRRLK
VLKTPILINKKGMPPFYVSSFCRWMSKRSIAR"
1562..1586
/gene="o137"
/note="terminator-like sequence"
complement(1585..3135)
/gene="f516"
complement(1585..3135)
/gene="f516"
/function="unknown"
/note="similar to Mg chelata subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=11
/protein_id="AAA67569.1"
/db_xref="GI:148173"
/translation="MHIMVTQGLMSLSIVHTRALGVNAPPIVEVHISKGLPGLTM
VGLPETTVEARDVRSAIINSYEPYPAKITINLAPADLPKGGRYDLPATALLAA
SEQLTANKLDEYELVGLALTGALRGVPGAISSEAIKSGRKIIIVAKDNEDEVGLIN
GEGHLIAHLQAVCAFLEGGKHALEXPKPTDAVSRAHQHLSDVIGQEQKRGLEITAA
GGHNLILPPTGKTMLASRINGLLPDLNNEEALESAAIILSVNAESVQKQWRQRP
RSPHSAISUTAMVGGAIIPGPEISLAHNGVLFDELPEFERITDALREPIESQIH
LSRTAKITYPARFQVLAAMNPSPTGHVQGNHNRCTPEQTLRYLNRLSGPFLDRFDLS
LEIPLPPGILSKTVVPGESSATVKORVMAERQFKRONKLNALWDSPEIRQFCKLE
SEDAWMEGLTILHLGLSIRAWQRLKLVARTIADIDQSDIITRQHLQEAUSYRAIDRL
IHQKLLT"
3287..3433
/note="predicted bend of 79 degrees"
3297..3326
/note="promoter matrix score of 74"
3336..12791
/note="corresponds to M10313 (1..9456)"
3336..3737
/note="corresponds to V00289 (1..402)"
3336..3676
/note="corresponds to M11651 (1..341)"
3379..5721
/note="V00290; ECILVX(1..2343)"
3458..3556
/gene="ilvL"
3458..3556
/gene="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGMDA operon leader peptide"
/protein_id="AAA67570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVVISVWVVIIPPCGAALGRGKA"
3696..4679
/gene="ilvG"
3696..4679
/gene="ilvG"
/note="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
```

```
/product="acetohydroxy acid synthase II, large subunit"
/protein_id="AAA67571.1"
/db_xref="GI:148175"
/translation="MNGAQVWVHALRAQGVNTVFGYPGGAIMPVVDALYDGGVEHLLC
RHEQGAAMAAIGYARATGKTGVCIAITSGPATNLITGLADALLDSIPVVAITQVSAP
FIGTDAFQEVLDVLSLACTKHSFLVQSLBELPRIMAEFADVACSGRPGPVLVDIPKD
IOLASGDLEPWFPTTVEVEVFPFAEVEQARQMLAKAOKPMLYVGGVGMAQAVPALRE
FLAATKMPATCTLKLGAVEADYPPYLGMLGMHGTKANFAVQECDDLIIAGARFDDR
VTGKLTNFAPHASVIHMDIDPAEMNKLQARVALQGLNALLPALQQLNQ"
4675..5340
/gene="o221"
4675..5340
/gene="o221"

Query Match      98.2%; Score 604; DB 1; Length 91414;
Best Local Similarity 99.8%; Pred. No. 1.3e-181;
Matches 615; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY. 1 ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCATTTAAACGCTGTCGCCA 60
Db 62161 ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCATTTAAACGCTGTCGCCA 62102

QY 61 GGCTCTGGTGCATCAACACTATGACCACTCGCTCAACACCGGTATC-CGGCCGGTGG 119
Db 62101 GGCTCTGGTGCATCAACACTATGACCACTCGCTCAACACCGGTATC-CGGCCGGTGG 62042

QY 120 CGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGCTGGTGGCGTGGGGT 179
Db 62041 CGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGCTGGTGGCGTGGGGT 61982

QY 180 TGGGACGCTATTTTCCCGCTCAGTGATTCGCTTTTGAAGTGTGGAAGTGGGACGGCGCGGC 239
Db 61981 TGGGACGCTATTTTCCCGCTCAGTGATTCGCTTTTGAAGTGTGGAAGTGGGACGGCGCGGC 61922

QY 240 TTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATC 299
Db 61921 TTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATC 61862

QY 300 GCTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTCAC 359
Db 61861 GCTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTCAC 61802

QY 360 CAATCCCAAAAAGTATTGTGTTTCTGCGCGCGCTATTTCGCAATTCATCATGCCGCAACA 419
Db 61801 CAATCCCAAAAAGTATTGTGTTTCTGCGCGCGCTATTTCGCAATTCATCATGCCGCAACA 61742

QY 420 GCCGCAACTGATGAGTATATCGTCTGCGCGCTCACCACCTATTGCTGCGATATTATTGT 479
Db 61741 GCCGCAACTGATGAGTATATCGTCTGCGCGCTCACCACCTATTGCTGCGATATTATTGT 61682

QY 480 GATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAGGACCAAGCA 539
Db 61681 GATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAGGACCAAGCA 61622

QY 540 GATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTGTTGGGAGCGCTGTAGC 599
Db 61621 GATGAAGCGCTGAATAAGATTTCGCGCTCGTTGTTTATGCTGTTGGGAGCGCTGTAGC 61562

QY 600 ATCGGCGAGGCATGCG 615
Db 61561 ATCGGCGAGGCATGCG 61546

RESULT 6
AE000458/c
LOCUS
DEFINITION Escherichia coli K12 MGL655 section 348 of 400 of the complete
genome.
ACCESSION AE000458 U00096
VERSION AE000458.1 GI:2367299
KEYWORDS
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
```

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

1 (bases 1 to 11509)  
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.P., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.  
The complete genome sequence of Escherichia coli K-12  
Science 277 (5331), 1453-1474 (1997)  
97426617  
9278503

2 (bases 1 to 11509)  
Blattner, F.R.  
Direct Submission  
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

3 (bases 1 to 11509)  
Blattner, F.R.  
Direct Submission  
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

4 (bases 1 to 11509)  
Plunkett, G. III.  
Direct Submission  
Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
On Sep 9, 1997 this sequence version replaced gi:1790254.  
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: markamber@gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

Location/Qualifiers  
1. .11509  
/organism="Escherichia coli K12"  
/mol\_type="genomic DNA"  
/strain="K12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:83333"  
81. .950  
/gene="pldA"  
/note="synonym: b3821"  
81. .950  
/gene="pldA"  
/EC\_number="3.1.1.32"  
/function="enzyme; Central intermediary metabolism: Phosphorus compounds"

FEATURES  
source

gene

CDS

/note="o289; 99 pct identical to PA1\_ECOLI SW: P00631; CGSCNo. 384"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane phospholipase A"  
/protein\_id="AAC76824.1"  
/db\_xref="GI:2367300"  
/translation="MRTIQGWLLPVFMLPMVAVYQAEATVKEVHDAPAVRGSIIANMLQ EHDNPFLLYPDTNYLIYTQFSDLNKEATASYDWAENARKDEVKFQSLAFPLWRGIL GPNVLGASYTKSWQLSNSESSSPFRETNYEPQLFLGPAFDYRFAGWTLRDVEMGY NHDNGRSDPTSRSNRLYTRLMAENGWLVVEVKPWYVVGNTDDNPDTIKYMGYQLK IGVLGDAVLGSAKQYNWNTGYGAEGLSYPIKTVRLYTVQVYSGYGESLIDYNFNQ TRVGVGVMLNDLF"  
965. .993  
/note="factor Sigma70; predicted +1 start at 4003392"  
1019. .1047  
/note="factor Sigma70; predicted +1 start at 4003446"  
1077. .2909  
/gene="recQ"  
/note="synonym: b3822"  
1077. .2909  
/gene="recQ"  
/EC\_number="3.6.1.-"  
/function="enzyme; DNA - replication, repair, restriction/modification"  
/note="o610; 99 pct identical to 607 amino acids of RECQ\_ECOLI SW: P15043 but has 3 additional N-ter residues"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent DNA helicase"  
/protein\_id="AAC76825.1"  
/db\_xref="GI:2367301"  
/translation="MNVAQAEVLNLESQAKQVLQETFGYQQFRPGQEIIIDTVLSGRD CLVMPGTGGKSLCYQIPALLNGLTVVVSPLISLMKDQVDLOQANGVAAACLNSTOT REQOLEVMTGCTGTQIRLLYIAPERLMDNFLEHLAHNPVLLAVDEAHGISQWGHDF RPEYAALGQLRQRPFTLPFMAITATADTTQDITVRLGLNDPLIQISSFDRPNIRYM LMEKFKPLDQLMRYVQEQRGSGIYYCNSRAKVEDTAARLQSKGISAAAYHAGLENNV RADYQEKFORDDLIIVATVAFGMGINKPNRVFVHFDPINIESYYQETGRAGRDGL PAEAMLFYDPADMAWLRRCLEKPGQQLQDIERHKLNAMGAFAEAOCTCRLVLLNYFG EGQEPGCGNCDICLDPKQYDGSIDQIALSTIGRVNQRFMGYVVEVIRGANNQRI R DYHDKLVYGVGRDKSHEHWVSIVIRQLIHLGLVTQNIACHSALQLTEARPVLAESS LQLAVPRIVALKPMQKSFQGNVDRKLFARLKLKRSIADESNNVPPYVVFNDATLIE MAEQMPITASEMLSVNGVGMKLERFGKPFMALIRAHVDGDDDEE"  
2104. .2140  
/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"  
3162. .3189  
/note="factor Sigma70; predicted +1 start at 4005588"  
3224. .3592  
/gene="yigJ"  
/note="synonym: b3823"  
3224. .3592  
/gene="yigJ"  
/function="orf; Unknown"  
/note="o122; sequence change shortens and changes N-terminus relative to earlier version (YIGJ\_ECOLI SW: P27846)"  
/codon\_start=1  
/transl\_table=11  
/product="orf, hypothetical protein"  
/protein\_id="AAC76826.1"  
/db\_xref="GI:2367302"  
/translation="MGYQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKA IIVFGSVFSLFVGDNVGTTARWGIIFALIIVETLAWFTTVASLFAFPQMERGYQLAKW IDGFAGALFAGFGIHLIISR"  
3612. .3645  
/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"  
complement(3654. .4070)  
/gene="yigK"  
/note="synonym: b3824"  
complement(3654. .4070)

repeat\_region

promoter

gene

CDS

repeat\_region

gene

CDS

/gene="yigK"  
/function="orf, Unknown"  
/note="f138; 100 pct identical to YIGK\_ECOLI SW: P27847"  
/codon\_start=1  
/transl\_table=11  
/product="orf, hypothetical protein"  
/protein\_id="AAC76827.1"  
/db\_xref="GI:1790258"  
/translation="MIAFEVLKWAAGAYLIWGIQWRAGAIIDLKSLASTQSRRLHF  
QRAVFNLTNPKSIVFLAALFPQIMPQQLMQYIVLGVTIVVDIIWMIGYATLAQ  
RIALWIKGPKQKALNKIFGSLFVLVGLLASARHA"  
4200..4230  
/note="factor Sigma70; predicted +1 start at 4006629"  
complement(4227..4254)  
/note="factor Sigma70; predicted +1 start at 4006612"  
4384..5406  
/gene="pldB"  
/note="synonym: b3825"  
4384..5406  
/gene="pldB"  
/EC\_number="3.1.1.5"  
/function="enzyme; Central intermediary metabolism:  
Phosphorus compounds"  
/note="o340; 99 pct identical amino acid sequence and  
equal length to PLDB\_ECOLI SW: P07000"  
/codon\_start=1  
/transl\_table=11  
/product="lysophospholipase L(2)"  
/protein\_id="AAC76828.1"  
/db\_xref="GI:2367303"  
/translation="MFQQQDWETRENAFAFTMGPLTFWRQRDEAFTGVDDIPVR  
FVFRQHQHDRVVICPGRIEYVKYAEIAYDLFHLGFDVLIIDHRQGRSGRLIADP  
HLGHVNFNDYDDLLAAFWQEVQPGPWRKRYILAHSGGAISTLFQRHPGVCDATA  
LTAPMFGIVIRMPSEFMARQILNWAHAHPRFRDGYAIGTGRWRALPFAINVLTHSRQRY  
RNRURFYADPTIRVGGPTVHWRESILAGEQVLGAGDADATPTLLQAEERVVQNR  
MHDRCFELRTAAGHPVEGGRPLVIKGAHYEILFEKDAMASVALHAIVDFFRHNSPSG  
NRSTEV"  
5381..5408  
/note="factor Sigma70; predicted +1 start at 4007807"  
5426..5455  
/note="factor Sigma70; predicted +1 start at 4007854"  
5438..5466  
/note="factor Sigma70; predicted +1 start at 4007865"  
5494..5523  
/note="factor Sigma70; predicted +1 start at 4007922"  
5526..6041  
Gene  
Query Match 96.7%; Score 595; DB 1; Length 11509;  
Best Local Similarity 98.9%; Pred. No. 7.2e-179;  
Matches 610; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
QY 1 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATCGATCAATTTAAACGCTGTGCGCA 60  
Db |||||  
4273 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATCGATCAATTTAAAGCCTGTGCGCA 4214  
QY |||||  
61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CCGCGCGGTG 118  
Db |||||  
4213 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTATCGCGCGCGGTG 4154  
QY |||||  
119 GCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTCATATTGCTGTTGCGGTGGGG 178  
Db |||||  
4153 GCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTCATATTGCTGTTGCGGTGGGG 4094  
QY |||||  
179 TTGGGACGCTATTTTCCCGCTCAGTGATTGCTTTGAAGTGTTGAAGTGGGAGCGCGG 238  
Db |||||  
4093 TTGGGACGCTATTTTCCCGCTCAGTGATTGCTTTGAAGTGTTGAAGTGGGAGCGCGG 4034  
QY |||||  
239 CTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTTGCAATTGACCTTAAT 298  
Db |||||  
4033 CTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTTGCAATTGACCTTAAT 3974  
QY |||||  
299 CGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGGAGTTTGTGTAATCTCA 358  
Db |||||

Db 3973 CGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGAGTTTTTGTGAATCTCA 3914  
QY 359 CCAATCCCAAAAGTATGTGTTTCTGGCGGCGTATTTCCGCAATTCATCATGCGCGCAAC 418  
Db |||||  
3913 CCAATCCCAAAAGTATGTGTTTCTGGCGGCGTATTTCCGCAATTCATCATGCGCGCAAC 3854  
QY 419 AGCCGCAACTGATGAGTATATCGTCTCGGCGTCAACCACTATGTGTCGATATTATG 478  
Db |||||  
3853 AGCCGCAACTGATGAGTATATCGTCTCGGCGTCAACCACTATGTGTCGATATTATG 3794  
QY 479 TGATGATCGGTTACGCCACCCCTTCTCAACCGGATGCTCTATGGATTAAAGACCAAAGC 538  
Db |||||  
3793 TGATGATCGGTTACGCCACCCCTTCTCAACCGGATGCTCTATGGATTAAAGACCAAAGC 3734  
QY 539 AGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTAG 598  
Db |||||  
3733 AGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTAG 3674  
QY 599 CATCGCGGAGGCATCGG 615  
Db |||||  
3673 CATCGCGGAGGCATCGG 3657

RESULT 7  
AB015396/c  
LOCUS  
DEFINITION Shigella flexneri 2a str. 301 section 359 of 412 of the complete genome. linear BCT 18-OCT-2002  
ACCESSION AE015396  
VERSION AE015396.1 GI:24054379  
KEYWORDS  
SOURCE Shigella flexneri 2a str. 301  
ORGANISM Shigella flexneri 2a str. 301  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.  
REFERENCE 1 (bases 1 to 10253)  
AUTHORS Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H., Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Kan, B., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L., Qiang, B.Q., Wen, Y.M., Hsu, Y.D. and Yu, J.  
TITLE Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157  
JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
PUBMED 12384590  
REFERENCE 2 (bases 1 to 10253)  
AUTHORS Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, Y., Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R. China

FEATURES  
Location/Qualifiers  
1..10253  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
complement(243..2167)  
/gene="yigE"  
/locus\_tag="SF3893"  
/pseudo  
complement(243..2167)  
/gene="yigE"  
/locus\_tag="SF3893"  
/note="similar to Escherichia coli K12 orf, conserved hypothetical protein gi: 1790248 (162 aa). BLAST with identity of 99% in 641 aa. This CDS has been disrupted by insertion. The sequence has been checked and is believed

to be correct."  
/pseudo  
/codon\_start=1  
/transl\_table=11  
472..1818  
/locus\_tag="SF3892"  
472..1818  
/locus\_tag="SF3892"  
/note="Residues 1 to 448 of 448 are 98 pct identical to  
residues 1 to 448 of a 448 aa protein from Shigella  
flexneri ref: NP\_085177.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45328.1"  
/db\_xref="GI:24054380"  
/translation="MPPDFFMHIGQALDLVSRVDSLRNPLTSLGDYLDPELISRCLAE  
SGTVTLKRRLPLEMNVWCIVGVLERKEPLHQIVNRLDIMLPGNRPFVAPSAVIQAR  
QRLGSEAVRRVFTKTAQLWNAATPHPHWCGLTLAIDGVFWRTPTDPTENDAAFPQOTH  
AGNPALYPQVKVVCQNELTSHLLTAAFGTMKNSENELAEQLIEQTGDNVTLTMDRGY  
YSLGLNGWSLAGERHWMIPLRKGAQYEEELRKLKGDHVLVKLKTSPQARKKWPGLGN  
EVTARLLTVTRKGKVCHELLTSMTDAMRFPGEEMADLYSNRWEIELGYREIKQTMQLSR  
LTLRSKPELVEQELMGVLLAYNLVRYQMIKWAHLKGYWPNQLSFSESCGMVMRLM  
TLQASPGRIPELMSDLASMGQLVKLPTRRGRAFPVRVKERPKYPTAPKKSQSA"  
2522..3487  
/gene="corA"  
/locus\_tag="SF3894"  
2522..3487  
/gene="corA"  
/locus\_tag="SF3894"  
/note="Residues 6 to 321 of 321 are 100 pct identical to  
residues 1 to 316 of a 316 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_290445.1"  
/codon\_start=1  
/transl\_table=11  
/product="Mg2+ transport, system I"  
/protein\_id="AAN45329.1"  
/db\_xref="GI:24054381"  
/translation="MGVPVMLSAPQLENNRLTRLEVESQPLVNAVWIDLVEPDDDER  
LRVQSLGSLATREPELIEASARFFEDDGLIHHSFFFEDEADHAGNSTVAFTR  
DGLRFLRLERLPAFLYRMARSQSMVDGNAYELLDDLLETKIEQLADEIENIYSDL  
EQLSRVIMEGHQDEYDEALSTLAELEDIGWKVRLCLMDTQRALNFLVRKARLPGGL  
EQAREILRDIESLPHNESLFQKVNFLMQAANGFINIEQNRILKIFSVSVSVFLPPTL  
VASSYGNFFPELPELKWVSGYPGAIIFMILAGLAPLYFKRKNL"  
complement(3528..3950)  
/gene="yigF"  
/locus\_tag="SF3895"  
complement(3528..3950)  
/gene="yigF"  
/locus\_tag="SF3895"  
/note="Residues 15 to 140 of 140 are 99 pct identical to  
residues 1 to 126 of a 126 aa protein from Escherichia  
coli K12 ref: NP\_418261.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45330.1"  
/db\_xref="GI:24054382"  
/translation="MHLLIHQNNSYGVFMSKEYMNDLSLSEKWKYRNFYDQHGFGPF  
WGATEYKAAFKALKVRQLTIQMNFIAFFCSWIYLFVLGLWKKAILVLLGLILSLFV  
GALIGVNIIGIAVAAAYAVANTNKWFYEKEVKGINTWSL"  
complement(3922..4362)  
/gene="yigG"  
/locus\_tag="SF3896"  
complement(3922..4362)  
/gene="yigG"  
/locus\_tag="SF3896"  
/note="Residues 9 to 146 of 146 are 95 pct identical to  
residues 1 to 138 of a 138 aa protein from Escherichia  
coli K12 ref: NP\_418262.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/protein\_id="AAN45331.1"  
/db\_xref="GI:24054383"  
/translation="MYESGHIAMLTGDSHKVDKFMRLMFIPTNSGKISRHRHYIFLFI  
INFIFAPLIIFNDGEAGFLVIVSTIALHYLVINMNCQRLRDSGFIYIKIYVFGTLAV  
YIISITWIAEDFACSGSGSMIFLICFYFSTFSLMLLAPTSSKQ"  
complement(4397..5287)  
/gene="rard"  
/locus\_tag="SF3897"  
complement(4397..5287)  
/gene="rard"  
/locus\_tag="SF3897"  
/note="Residues 1 to 296 of 296 are 99 pct identical to  
residues 1 to 296 of a 296 aa protein RARD\_ECOLI sp:  
P27844"  
/codon\_start=1  
/transl\_table=11  
/product="Rard protein"  
/protein\_id="AAN45332.1"  
/db\_xref="GI:24054384"  
/translation="MDAKQTRQGVLLALAAAYFIWGIAPAYFKLIYVYPADIELTHRVI  
WSFFFMVLMISICROWSYKTLIQTPQKIFMLAVSAVLIGGNWLLFIWAVNNHMLEA  
SLGYFINPLVNIVLGMIFLGERFRRNQWLAVILAICGVLVQLWTFGSLPIIALGLAFS  
FAFYGLVRKKIAVEAQTMLIETMWLLPVAIYLFIAIDSSSTSHMGQNPMSLNLLLIA  
AGIVTVPLLCFTAAATRLRLSTLGGFYIIGPTLMFLAVIFYGEKPGADKMVTFTFI  
WVALAIFVMDAIYQRTSK"  
complement(5339..5824)  
/gene="yigI"  
/locus\_tag="SF3898"  
complement(5339..5824)  
/gene="yigI"  
/locus\_tag="SF3898"  
/note="Residues 1 to 161 of 161 are 100 pct identical to  
residues 1 to 161 of a 161 aa protein from Escherichia  
coli O157:H7 ref: NP\_312777.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45333.1"  
/db\_xref="GI:24054385"  
/translation="MKCKDDMSAVLTAEQALKLVGMFVYHMPFNALGMELERYEKE  
FAQLAFNQPMVMGVNWAQSIHGGVIAALDVAGLVCVGSTLTRHETISEDELQRL  
SRMGTTDLRVLYLRPGRGERFTATSSLLRAGNKVAVARVELHNEEQLYIASATATMV  
G"  
5971..6840  
/gene="pldA"  
/locus\_tag="SF3899"  
5971..6840  
/gene="pldA"  
/locus\_tag="SF3899"  
/note="Residues 1 to 289 of 289 are 100 pct identical to  
residues 1 to 289 of a 289 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_290453.1"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane phospholipase A"  
/protein\_id="AAN45334.1"  
/db\_xref="GI:24054386"  
/translation="MRTLQGWLLPVFMPLPMVAYQEAATVKEVHDAPAVRGSIIANMLQ  
EHDNPFTLYPYDTNYLIYQTSDLNKEAIASYDWAENARKDEVKFQLSLAPFLWRGIL  
GPNVLGASYTQKSWQLSNSSESSPFRETNYEPQLFLGFATDYRFAGWTLRDVEMGY  
NHDSNGRSDPTSRSNRNLRYTRLMAENGWLVEVKPWYVVGNTDDNPDITKVMGYQLK  
IGYHLGDAVLSAKGQYNWNTGYGGAELGLSYPIITKHVRLYITQVSYSGYESLIDYFNQ  
TRVGVGWLNLDLF"  
5988..8802  
/gene="recQ"  
/locus\_tag="SF3900"  
5988..8802  
/gene="recQ"  
/locus\_tag="SF3900"  
/note="Residues 1 to 604 of 604 are 99 pct identical to  
residues 8 to 611 of a 611 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_290454.1"  
/codon\_start=1

gene

CDS

gene

CDS

gene

CDS

gene

CDS

Query Match 93.9%; Score 577.6; DB 1; Length 10253; Best Local Similarity 98.1%; Pred. No. 2.7e-173; Matches 606; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

1 ATGACCTTAGAATGGTGGTTTGCCCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 60  
10168 ATGACCTTAGAATGGTGGTTTGCCCTACCTGCTGACATCGATCATTTTAAGCCTGCGCCA 10109

61 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCAACCGGTTAT--CCGCGCGGTG 118  
10108 GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACCAACCGGTTATCGCGCGCGGTG 10049

119 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTTGGCTGGGG 178  
10048 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTTGGCTGGGG 9989

179 TT-GGGACGCTATTTCCCGCTCAGTGATGCGTTTGAAGTGTGGAAGTGGGCAAGCGCG 237  
9988 TTGGGGACGCTATTTTCCCGCTCAGTGATGCGTTTGAAGTGTGGAAGTGGGCAAGCGCG 9929

238 GCTTACTTGAATTGGCTGGGAATCCAGAGTGCGCGCGCTGGTGAATGACCTTAAA 297  
9928 GCTTACTTGAATTGGCTGGGAATCCAGAGTGCGCGCGCTGGCGCAATTGATCTTAAA 9869

298 TCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 357  
9868 TCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 9809

358 ACCAATCCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGGCAATTCATCATCGCGCAA 417  
9808 ACCAATCCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGGCAATTCATCATCGCGCAA 9749

418 CAGCCGCAACTGATGCAGTATATCGTGTGCGCGTCACCACCTATTGTGGTCGATATTAT 477  
9748 CAGCCGCAACTGATGCAGTATATCGTGTGCGCGTCACCACCTATTGTGGTCGATATTAT 9689

478 GTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG 537  
9688 GTGATGATCGGTTACGCCACCCTTGCTCAAGGATTGCTCTATGGATTAAAGGACCAAG 9629

538 CAGATGAAGGCGCTGAATAAGATTTTTCGCTCGTTGTTTATGCTGTGGGAGCGCTGTTA 597  
9628 CAGATGAAGGCGCTGAATAAGATTTTTCGCTCGTTGTTTATGCTGTGGGAGCGCTGTTA 9569

598 GCATCGCGGAGGCATGCG 615  
9568 GCATCGCGGAGGCATGCG 9551

RESULT 8  
AE016990  
LOCUS Shigella flexneri 2a str. 2457T section 13 of 16 of the complete genome. BCT 22-APR-2003  
DEFINITION Shigella flexneri 2a str. 2457T  
ACCESSION AE016990  
VERSION AE016990.1 GI:30042943  
KEYWORDS Shigella flexneri 2a str. 2457T  
SOURCE Shigella flexneri 2a str. 2457T  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.  
REFERENCE 1 (bases 1 to 290976)  
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.

TITLE Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T  
JOURNAL Infect. Immun. 71 (5), 2775-2786 (2003)  
PUBMED 12704152  
REFERENCE 2 (bases 1 to 290976)  
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.  
Direct Submission  
Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source Location/Qualifiers  
1..290976  
/organism="Shigella flexneri 2a str. 2457T"  
/mol\_type="genomic DNA"  
/strain="2457T"  
/serotype="2a"  
/db\_xref="taxon:198215"  
complement(68..1405)  
/gene="yjdE"  
/locus tag="S3622"  
complement(68..1405)  
/gene="yjdE"  
/locus tag="S3622"  
/function="putative transport; Not classified"  
/note="residues 1 to 445 of 445 are 92.35 pct identical to residues 1 to 445 of 445 from Escherichia coli K-12: B4115"  
/codon start=1  
/transl\_table=11  
/product="putative amino acid/amine transport protein"  
/protein\_id="AAP18666.1"  
/db\_xref="GI:30042944"  
/translation="MSSDADAHKVLIPVTLVSGNIMSGVFLLPANLASTGGIAIY  
GWLVTIGALGLSMYAKMSFLDPSPGSYAYARRCFGLGQTNVLYWLAQWIGNI  
AMVIVGVLSYFFPILKDLPLVLTITCVVAVLWIFVLLNIVGPKWITRVQAVATVLALI  
PIVIAVFGWFWFRGETYMAAMVNSGLGTGFGAIQSTLNVTLSFVIGVESASVAGVVK  
NPKRNVPIATIGGVLLIAAVCYVLSITAIMGMI PNAALRVASPFDAARMALGDTAGA  
IVFCAAAGCLGSLGWTLLAGTAKAADDGFPPIFARVNAKAGTPVAGLIIVGILM  
TIFQLSSI SPNATKFEGLVSSVSFTLVPLVYTC AALLLLGHGF GKARPAYLAVTT  
IAFLYCIWAVVSGAKEVMWSFVTLMVITAMVALYNRLHKNPYPLDAPISKD"  
complement(1542..2303)  
/gene="adiY"  
/locus tag="S3623"  
complement(1542..2303)  
/gene="adiY"  
/locus tag="S3623"  
/function="putative regulator; Not classified"  
/note="residues 1 to 253 of 253 are 93.28 pct identical to residues 1 to 253 of 253 from Escherichia coli K-12: B4116"  
/codon start=1  
/transl\_table=11  
/product="putative ARAC-type regulatory protein"  
/protein\_id="AAP18667.1"  
/db\_xref="GI:30042945"  
/translation="MRICSDQPCIVLLTEKDVWIRVNGKEPISLKHMHMALLNCENNI  
IDVSSLNNTLVAHISHDIIKDYIRFLNKLQSLQIPVWQSRATPLTLPCLTPDVERVAA  
QHSMMPAETESSEKERTALLFTVLSRFLDSKKFLSLMVMYMLNVCSDSVYQIIESDIH  
KDWNLMSVASCCLSPSLKKLKSNTSYSQIITTCRMRYAVNLMMDGKNISQVSDQ  
SCGVNSTSYFISVFKDFYGMTPLHYVSQHRERTVA"  
complement(2628..4895)  
/gene="adiA"  
/locus tag="S3624"  
complement(2628..4895)  
/gene="adiA"  
/locus tag="S3624"  
/function="enzyme; Degradation of small molecules: Amino acids"  
/note="residues 1 to 755 of 755 are 99.47 pct identical to residues 2 to 756 of 756 from Escherichia coli K-12: B4117"



```
/codon_start=1
/transl_table=11
/product="biodegradative arginine decarboxylase"
/protein_id="AAP18668.1"
/db_xref="GI:30042946"
/translation="MKVLIVSEFLHQDTWVGNAVERLADALSQQNVTVIKSTSFDDG
FAILSNEAIDCLMFYSQMEHDEHQNVRLQIKLHERQNVVPVILGDRKALAAAD
RDLELVDEFWILEDTADFIAGRAVAAMTRYRQQLLPFLFSALMKYSDIHEYSWAAP
GHQGVGFTKTPAGRFYHDYGENLFRITDMGIERITSLGSLDHTGARGSEKYYAARVF
GADRSWSVVGTSGNSRNTIMQACMTDNDVVVDNRNCHKSIEQGLMLTGAKPVVMVPSR
NRYGIGIYPOEMQPETLQKKISEPLTKDKAGQKPSYCVVNTCTYDGVCCYNAKEAQ
DLLEKTSRDLHDFEAWGYARENPDIYADHYAMRGEPGDNHNGPTVPFATHSIHKLNLALS
QASVTHVREGRAINFNFNOAYMMHATTSPLYAICASNDVAVSMMDCNSGLSLTQEV
IDEAVDFRQAMARLYKEFTADGSWFFKPNWKEVVDTPQTKGTIDFADAPTCLLTIVQD
CWNHPGESWHGFKDIPDNWSMLDPIKVSITLAPMGDEGELEETGVPAALVTAWLGRH
GIVPTRTDFQIMFLFSMGVTRKGWGLVNTLCSFKHYDANTPLAQWMPBELVEQYPD
TYANMGIDHLDGTMFAWLKENNPGARLNEAYSGLPVABITPREAYNAIVDNNVELYSI
ENLPGRIAANSVIPYPPGIPMLLSGENFGCKNSPQVSILRSLSQSWDHHFRFEHETEG
TEIIDGIYHVMCVKA"
complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/note="residues 1 to 302 of 302 are 99.66 pct identical to
residues 1 to 302 of 302 from Escherichia coli K-12 :
B4118"
/codon_start=1
/transl_table=11
/product="regulator of melibiose operon"
/protein_id="AAP18669.1"
/db_xref="GI:30042947"
/translation="MNTDTFMCSDSEKQTRSPLSLYSEYQRMIEFRAPHIMPTSHWH
GOVENVPFDGVEYLINNEKYNINQGHITLFWACTPHQLTDTGTCSMAIFNLPHLH
FLSWPLDKLINHTHGMVIVKSLATQQLSPFEVRRWQQSLNSPNEIQRQLAIDEIGLM
LKRESLSSWEPILVNKTSRTHKNSVSRHAQPVVSQMLGFAENYDQALTINDVAEHVK
LNANYAMGIFQVYNQLTMKQYITAMRINHVRALLSDTDXSILDIALTAGFRSSRFPYS
TFGKIVGMSPPQYRKLSQRRQTFPG"
6060..6114
/note="BOXC-20"
6285..7640
/gene="mela"
/locus_tag="S3626"
6285..7640
/gene="mela"
/locus_tag="S3626"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="residues 1 to 451 of 451 are 98.89 pct identical to
residues 1 to 451 of 451 from Escherichia coli K-12 :
B4119"
/codon_start=1
/transl_table=11
/product="alpha-galactosidase"
/protein_id="AAP18670.1"
/db_xref="GI:30042948"
/translation="WMSAPKTIFIGAGSTIFVKNILGDFVHREALKTAHIALMDIDPT
RLESHIVVRKLMDSAGASGKITCHTQQKEALEDAFVVFVAFQIGGYEPCVTVDFEVC
KRHGLEQTIADTLPGGGIMRALRTIPLHWOICENMTEVCPCATMLNYPNPMAMNTWAM
YARYPIKQVGLCHSVQGTAEELARDLNI DPATLRYRCAGINHMFAFYLELERKTIADGS
YVNLYPELLAAEYEAQAPKPNTHGNTRCNQIVRYEMFKILGYFVFTESSEHFAEYTPWF
IKLGRDLIERIKVPLDEYPRKVEQLANWHELEEYKNASRIDIKPREYASTINMA
INTGSPSVIYGNVRNDGLDNLDPQGCVEVACLVDANGIOPTKVGTLPFSLAALMOTN
INVQTLTEAILTENRDRVYHAMNDPHTAAVLGIDEIYALVDDLLIAHGDWLPFSLH
R"
7755..9164
/gene="melB"
/locus_tag="S3627"
7755..9164
/gene="melB"
```

```
/locus_tag="S3627"
/function="transport; Transport of small molecules:
Carbohydrates, organic acids, alcohols"
/note="residues 1 to 469 of 469 are 96.37 pct identical to
residues 1 to 469 of 469 from Escherichia coli K-12 :
B4120"
/codon_start=1
/transl_table=11
/product="melibiose permease II"
/protein_id="AAP18671.1"
/db_xref="GI:30042949"
/translation="MTTKLSYGFAGFKDFAGICIVMYLMYYTVDVGLSVGLVGTLF
LVARIWDAINDPIMGIWVNAIATSRWKGFKPWILIGTLANSVILFLFSAHLFEGTTQI
FVCVTYILLWGMTYTIMDIPFWSLVPTITLDKREQLVPYPRPFASLAGFVTAGVTL
PFVNYVGGDRGFQFTLVLIATFIVITITLRNVHEVFSDDNQSSAEGSHLTAKA
IVALIYKNDQLSCLLGMALAYNVASNIITGPAIYFYSVIGDADLFPYLYSAGAANL
VTLVFPRLVKLSRRLWAGASILPVLSCGVLMLMALSYHNVLIVIAIGILLNVGT
ALFWVLQVIMVADTVDYGEYKLVHRCESIAYSVQTMVVKGSFAFAFFIAVILMIGY
VPNEQSTQALLGMQFIMIALPTLFFFMVTLILYFRFYRLNGDTRRIQIHLLDKYRKV
PPEPVHADIPVGAVSVDKA"
9250..9283
/note="REP (repetitive extragenic palindromic) element,
contains 1 REP sequence"
complement(9303..9932)
/gene="yjdF"
/locus_tag="S3628"
complement(9303..9932)
/gene="yjdF"
/locus_tag="S3628"
/note="residues 1 to 209 of 209 are 83.25 pct identical to
residues 1 to 209 of 209 from Escherichia coli K-12 :
B4121"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP18672.1"
/db_xref="GI:30042950"
/translation="MTRTLKPLIINTGALALTLILIYTGISAHDKLTWLEVTPIIV
VPLLATARRYPLTLLYLIFFHAILLMVGQYTYAKVPVGEVQEWLGLSRNPYDK

Query Match      93.9%;   Score 577.6;   DB 1;   Length 290976;
Best Local Similarity 98.1%;   Pred. No. 4.7e-173;
Matches 606;   Conservative 0;   Mismatches 9;   Indels 3;   Gaps 2;

Qy      1  ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 60
Db      251870  ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAAAGCCTGTCGCCA 251929

Qy      61  GGCTCTGGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CCGGCCGGGTG 118
Db      251930  GGCTCTGGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTATCGCGCGCGGTG 251989

Qy      119  GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTATATTTGCTGCTGGTGGCGTGGGG 178
Db      251990  GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTATATTTGCTGCTGGTGGCGTGGGG 252049

Qy      179  TT-GGGACGCTATTTTCCCGCTCAGTGATTGCGGTTTGAAGTGTGAAGTGGCGAGCGCG 237
Db      252050  TTGGGACGCTATTTTCCCGCTCAGTGATTGCGGTTTGAAGTGTGAAGTGGCGAGCGCG 252109

Qy      238  GCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAAA 297
Db      252110  GCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCGCTGGTGCAATTGATCTTAAA 252169

Qy      298  TCGCTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 357
Db      252170  TCGCTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 252229

Qy      358  ACCAATCCCAAAAGATATTGTGTTTCTGGCGGGCGCTATTTTCGGCAATTCATCATGCCGCAA 417
Db      252230  ACCAATCCCAAAAGATATTGTGTTTCTGGCGGGCGCTATTTTCGGCAATTCATCATGCCGCAA 252289

Qy      418  CAGCCGCAACTGATGATGATATATCGTGTGCTCGGCGGTCAACCACTATTGTGTCGATATTATT 477
```

Db	252290	CAGCCGCAACTGATGTCAGTATATCGTCTCGGCGTCACCACTATTGTGTCGATATTATT	252349
QY	478	GTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAAG	537
Db	252350	GTGATGATCGGTTACGCCACCTTGCTCAAGGATTGCTCTATGGATTAAAGGACCAAAG	252409
QY	538	CAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA	597
Db	252410	CAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA	252469
QY	598	GCATCGCGGAGGCATGCG	615
Db	252470	GCATCGCGGAGGCATGCG	252487
RESULT 9	AP002567/c		
LOCUS	Escherichia coli O157:H7	318703 bp	DNA linear BCT 07-MAR-2001
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 18/20.		
ACCESSION	AP002567	BA000007	
VERSION	AP002567.1	GI:13364198	
KEYWORDS			
SOURCE	Escherichia coli O157:H7		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (sites)		
AUTHORS	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.		
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak		
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)		
MEDLINE	20198780		
PUBMED	10734605		
REFERENCE	2 (sites)		
AUTHORS	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.		
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655		
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		
MEDLINE	20557356		
PUBMED	11108008		
REFERENCE	3 (sites)		
AUTHORS	Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.		
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
JOURNAL	Gene 258 (1-2), 127-139 (2000)		
MEDLINE	20564182		
PUBMED	1111050		
REFERENCE	4 (sites)		
AUTHORS	Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.		
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12		
JOURNAL	DNA Res. 8 (1), 11-22 (2001)		
MEDLINE	21156231		
PUBMED	11258796		
REFERENCE	5 (bases 1 to 318703)		
AUTHORS	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.		

TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)		
COMMENT	genome project.		
FEATURES	Location/Qualifiers		
source	1..318703		
	/organism="Escherichia coli O157:H7"		
	/mol_type="genomic DNA"		
	/strain="O157:H7"		
	/sub_strain="RMD 0509952"		
	/db_xref="taxon:83334"		
gene	103..648		
	/gene="ECs4723"		
CDS	103..648		
	/gene="ECs4723"		
	/note="similar to WECD_ECOLI gi 1790225 percent identity 98 in 181 aa (Conserved in E.coli K-12)"		
	/codon_start=1		
	/evidence=not experimental		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="BAB38146.1"		
	/db_xref="GI:13364199"		
	/translation="MQAKIAASNTGELDALQQLGFSLVEGEVDLALPVNNASDGAUV AQETDIPALQLASAPAFQSRFRAPWYAPDASSRFYAQWIENAVRGTFDHOCLILRAA SGDIRGVSLRELNATDARIGLLAGRGAGAEMLQALNWAYARGKTTLRVATQMGNTA ALKRYIQSGANVESTAYWLYR"		
gene	653..1783		
	/gene="ECs4724"		
CDS	653..1783		
	/gene="ECs4724"		
	/note="similar to WECE_ECOLI gi 2367285 percent identity 99 in 376 aa (Conserved in E.coli K-12)"		
	/codon_start=1		
	/evidence=not experimental		
	/transl_table=11		
	/product="putative regulator"		
	/protein_id="BAB38147.1"		
	/db_xref="GI:13364200"		
	/translation="MIPFNAPPVVGTELDYMQSAMSGKLCGPGGFTRRCCQWLEQRF GSAKLLTPSCSTLEMAALLDIQPGDEVIMPSYTFVSTANAFVLRGAKIVFVDVRF DTWNIDETLIEAATDKTRVIVPVHAYAGVACEMDTIMALAKKHNLFVEDAAQGVMT YKRALGTIGHIGCFSFHETKNYTAGGEGATLNDKALIERAEIIREKGTNRSQFFR GQVDKYTRWDIGSSYLMSDIQAAYLWAQLEAADRIHQORLALWQNYDALAPLAKAGR IELPSIPDGCLQNAHMFYIKGRDIGDRSALINFLKBAEIMAVPHYIPLHGCAPAGERFG EFHGEDRYTTKESERLLRLPLFYNLSPVNQRTVIATLLNYS"		
gene	1785..3035		
	/gene="ECs4725"		
CDS	1785..3035		
	/gene="ECs4725"		
	/note="similar to WZXE_ECOLI gi 1790227 percent identity 98 in 416 aa (Conserved in E.coli K-12)"		
	/codon_start=1		
	/evidence=not experimental		
	/transl_table=11		
	/product="putative cytochrome"		
	/protein_id="BAB38148.1"		
	/db_xref="GI:13364201"		
	/translation="MSLAKASLWTAASLTVKIGALLVGKLLAVSFGPAGLGIAANFR QLITVLGVLGAGIFNGVTKYAQYHDNPQQLRVVGTVSSAMVLGFTLMLVFLVLA APTSLGLFGNHDYQGLVRLVALVQMGIANGNLLALMKGFRDAAGNALSLIVGSLIGV LAYVYSYRLGGYEGALLGALIPALVVIIPAAVNLIKRGAIPLSYLKPSWDNGLAGQLS KFTLMTLTJTSVTLPVAYIMMRKLLAAQYSWDEVGIMQGVSSISDAYLQFITASFVYL LPTLSRLTEKRDITREVVKSLKPVLPVAAAASFVWLLRDRFAIWLILLSNKFTAMRDLF AWQLVDVLKVGAVYFGLVIAKASLRFYILAEVVSQFTLLMVFAHWLIPTHGALGAAQ AYMATYIVYFSLCCGVFLLWRRRA"		
gene	3032..4111		
	/gene="ECs4726"		
CDS	3032..4111		
	/gene="ECs4726"		

/note="N-terminal part (1-62 in 359 aa) is similar to  
YIFM\_1 ECOLI gi|2367286 percent identity 96 in 62 aa,  
C-terminal part (156-359 in 359 aa) is similar to  
YIFM\_2 ECOLI gi|2367287 percent identity 96 in 204 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB38149.1"  
/db\_xref="GI:13364202"  
/translation="MTVLIHVLGSDIPHNRTVLRFFNDALAAATSEHAREBFMVVGKDD  
GLSDCPALSVQFPFGKSLAEAVIAKAKANRQORFFHGFQFNPTLWALLSGGKIPS  
QFYWHIWGADLYELSSGLRYKLFYLRRLAQKRVGCVFATRGDLSFFAKTHPKRGEL  
LYFPTMDASLNTMANDRQREGKMTILVNGSDRSNEHIAALRAVHOQFGDTVKVVP  
MGYPHNEAYIEVRQAGLELSEENLQVLEKLEFDAYTLRLQCALGYFIFARQOG  
IGTCLLIQAGIPCVLNRENPFQDMTEQHLFVLFTTDDLNEIDIVREARQLASVDKN  
TIAFSPNYLQWQORALIAAGEVA"  
4108. .5460  
/gene="ECs4727"  
4108. .5460  
/gene="ECs4727"  
/note="similar to WECF ECOLI gi|2367288 percent identity  
99 in 450 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="TDP-Fuc4NAC:lipidII transferase"  
/protein\_id="BAB38150.1"  
/db\_xref="GI:13364203"  
/translation="MSLLQFSGLFVFWLLCTLFATLTWFEPRFRVRNFNFVFSLLPL  
LTFFGFPLTSLVFRFDVGVAPEILLQALLSAGCFYAVYVTVYKRLRKRVADVPR  
RPLFTMNRVETNLTWILMGIALVSIGIFMHNGLFLRLNSYSQIFSSSEVSGVALKR  
FFYFFPAMLVVFLRQDSKAMFLFLVSTVAFGLLTVMIVGGTRANIIIAFAIFLFIG  
IIRGWSLMLAAAGVLGIVGMFWLALKRYGMNVSGDEARYFLYLTRDFTSPWENLA  
LLQNYDNIDFQGLAPIVRDFYVFIPLSPWLPGRSPMSVNSANYFTWEVLNNHSGLAIS  
PTLISLVMGGALFIPLGAIIVGLIIKKFWDWLYELGNRETNRYKAAIILHSFCFGAIF  
NMIVLAREGLDSFVSRVVFIVVFGACIMIAKLILYLFESAGLIHKRTKSSLRTQVEG  
"  
5463. .6203  
/gene="ECs4728"  
5463. .6203  
/gene="ECs4728"  
/note="similar to WECG ECOLI gi|2367289 percent identity  
98 in 246 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="probable UDP-N-acetyl-D-mannosaminuronic acid  
transferase"  
/protein\_id="BAB38151.1"  
/db\_xref="GI:13364204"  
/translation="MNNNTTAPTPTLRLGLQLIGWRDMQHALDYLFADGOLKQGTLVAI  
NAERMLTIEENAEVRELINAEFKYADGISVVRSVRKKYPQAQVSRVAGADLWHEELMA  
RAGKGTTVFLVGKPEVLQAQTEAKLRNQMNWNIIVGSDGYFKPEQROALFERIHASG  
AQIVTVAMGSPKQIFMRDCRLVHPDALYMGVGGTYDVFTHGVKRAPKIWTGLLEWL  
YRLSQPSRIKRLQLRLYLRWHTGNL"  
6394. .7779  
/gene="ECs4729"  
6394. .7779  
/gene="ECs4729"  
/note="similar to YIFK ECOLI gi|2367290 percent identity  
98 in 461 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative amino acid/amine transport protein"  
/protein\_id="BAB38152.1"  
/db\_xref="GI:13364205"  
/translation="MADNKPQLQGLREARHIELIALGGTIGVGLFMGAATLKWAGPS  
VLLAYIAGLFVFFIMRSMGEMLEFPVTGSAFYAHRVMSPPFFGYLTAWSYWFMWMA  
VGISEITAGVYQVFPFEMAQWIPALIAVALANLAARLYLGEIEFDFWAMIKVTT  
IIVMIVIGLGVIFFGFNGGQSIGSFNSLTHEGFFAGGKWKGLTALCIVVASYQGVEL

Query Match 92.9%; Score 571.2; DB 1; Length 318703;  
Best Local Similarity 97.4%; Pred. No. 5.4e-171;  
Matches 602; Conservative 0; Mismatches 13; Indels 3; Gaps 2;  
Qy 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAAACGCTGTGCGCA 60  
Db 37470 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCAATCATTTTAAAGCCTGTGCGCA 37411  
Qy 61 GGCTCTGGTGCAATCAACACTATGACACCTCGCTCAACACGCTTAT--CCGGCCGGTG 118  
Db 37410 GGCTCTGGTGCAATCAACACTATGACACCTCGCTCAACACGCTTATCGCGCGCGGTG 37351  
Qy 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGCGTGGGG 178  
Db 37350 GCGTCTATTGCGGGCTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGCGTGGGG 37291  
Qy 179 TT-GGGACGCTATTTCCTCCGCTCAGTGATTGCGGTTTGAAGTGTGAAGTGGGCAGGCGG 237  
Db 37290 TTGGGACGCTATTTCCTCCGCTCAGTGATTGCGGTTTGAAGTGTGAAGTGGGCAGGCGG 37231  
Qy 238 GCTTACTTGTATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAA 297  
Db 37230 GCTTACTTGTATTGGCTGGGAATACAGCAATGGCGCGCGCTGGTGCAATTGATCTTAA 37171  
Qy 298 TCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 357  
Db 37170 TCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 37111  
Qy 358 ACCAATCCCAAAAGTATTGTTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCGGCAA 417  
Db 37110 ACCAATCCCAAAAGTATTGTTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCGGCAA 37051  
Qy 418 CAGCGCAACTGATGTCAGTATATCGTCTCGCGCGTCACCACTATTGTGTCGATATTATT 477  
Db 37050 CAGCGCAACTGATGTCAGTATATCGTCTCGCGCGTCACCACTATTGTGTCGATATTATT 36991  
Qy 478 GTGATGATCGGTTACGCCACCCTTGCTCAACGGATTTGCTATGGATTAAAGGACCAAG 537  
Db 36990 GTGATGATCGGTTACGCCACCCTTGCTCAACGGATTTGCTATGGATTAAAGGACCAAG 36931  
Qy 538 CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTA 597  
Db 36930 CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTA 36871  
Qy 598 GCATCGCGGAGGCATGCG 615  
Db 36870 GCATCGCGGAGGCATGCG 36853  
RESULT 10  
AE016769/c 301566 bp DNA linear BCT 24-DEC-2002  
LOCUS Escherichia coli CFT073 section 15 of 18 of the complete genome.  
DEFINITION Escherichia coli CFT073 section 15 of 18 of the complete genome.  
ACCESSION AE016769 AE014075  
VERSION AE016769.1 GI:26110701  
KEYWORDS  
SOURCE Escherichia coli CFT073  
ORGANISM Escherichia coli CFT073  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 (bases 1 to 301566)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.  
Extensive Mosaic Structure Revealed by the Complete Genome Sequence  
of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
12471157  
JOURNAL  
PUBMED  
REFERENCE  
2 (bases 1 to 301566)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,

Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,P.R.  
Direct Submission  
Submitted (20-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES  
source

1. 301566  
/organism="Escherichia coli CFT073"  
/mol\_type="genomic DNA"  
/strain="CFT073"  
/db\_xref="taxon:199310"  
complement(72..1100)  
/gene="waaW"  
/locus\_tag="c4450"  
complement(72..1100)  
/gene="waaW"  
/locus\_tag="c4450"  
/note="Residues 1 to 341 of 342 are 98.24 pct identical to  
residues 1 to 341 of 341 from GenPept.129 :  
>gb|AAC69673.1| (AF019746) UDP-galactose:(galactosyl) LPS  
alpha1,2-galactosyltransferase WaaW [Escherichia coli]"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-galactose:(galactosyl) LPS  
alpha1,2-galactosyltransferase"  
/protein\_id="AAN82886.1"  
/db\_xref="GI:26110702"  
/translation="MDLLAESITEVAVSGRIANTDRVLNIAYGIDRNLFGAAVSMQS  
VVMHNPDLAVKXFLHFDYIDEDYLQRVNAFTSKNAVVEIRIYKVSSAFIDIFPSLKQW  
SYATFFRLVAFQYLSETIENLIYIDADYICKGSLAGLIDINFQDKFAAIVKDVPFMQ  
EKPAKRLAIEGLPCGNFYFAGVYVYLOLEAWAKNDFMNAIAMLASDPQHTKYKCLDQDI  
LNLFSSHCFITSGDYIDCFYGDYELKNKSDDEDYKRTITDDTKLIHYGVGTVKPNWDMT  
NYPQKYFNEAYQVSCWNDVAFIPATNEKQYQVKYQHAKNGDTFNAFIYFIKFKLNK  
YRKLFQGQ"  
complement(1126..1818)  
/gene="rfay"  
/locus\_tag="c4451"  
complement(1126..1818)  
/gene="rfay"  
/locus\_tag="c4451"  
/function="enzyme; Macromolecule metabolism:  
Lipopolysaccharide"  
/note="Escherichia coli K-12 ortholog: b3625; Escherichia  
coli O157:H7 ortholog: z5052"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide core biosynthesis protein  
rfay"  
/protein\_id="AAN82887.1"  
/db\_xref="GI:26110703"  
/translation="MITSIRYRGFSFYKQNDNKKYKEIFDEILAYNFKTVKVLNRNIDD  
TKVSLDITKYGRYVFKVFAKTKRNERFLKSFVKGYQVQNLIIVETDRVRSAGLTFPND  
FYFLAERKIFNYASVFIMLIYEYVEGVELNDMPPIIPEDYKAEIKASKEKLHALNMLSGD  
PHRGNFIVSKDGVRIIDLSGKSCTAERKARDRLAMERHLGIANEIKDYGYYSVIYRTK  
LRXFIKKLKGKA"  
complement(1828..2823)  
/gene="rfaj"  
/locus\_tag="c4452"  
complement(1828..2823)  
/gene="rfaj"  
/locus\_tag="c4452"  
/EC\_number="2.4.1.58"  
/function="enzyme; Surface polysaccharides and antigens"  
/note="Escherichia coli K-12 ortholog: b3626; Escherichia  
coli O157:H7 ortholog: z5051"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide 1,2-glucosyltransferase"  
/protein\_id="AAN82888.1"  
/db\_xref="GI:26110704"

gene

CDS

AFNMNNIYLAQFDFTIYTLKNELHDSHRKYQQTITDKTVLIHYTGITKPMHSHWAGYP  
SASYFNIAREQSPWKYPPLKEARTVAEMQKQYKHLFAHGEYIKGITSLIKYLKK"  
complement(2840..3856)  
/gene="rfai"  
/locus\_tag="c4453"  
complement(2840..3856)  
/gene="rfai"  
/locus\_tag="c4453"  
/EC\_number="2.4.1.44"  
/function="enzyme; Surface polysaccharides and antigens"  
/note="Escherichia coli K-12 ortholog: b3627; Escherichia  
coli O157:H7 ortholog: z5053"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide 1,3-galactosyltransferase"  
/protein\_id="AAN82889.1"  
/db\_xref="GI:26110705"  
/translation="MSAHYFNPQEMINKTIIIFDERPAVSVASSFHVAYGIDQNFLLFGC  
GVSITSVLLHNDVSVFVHFVIDDIPEDIQRLSQLAKSYHTCIQIHLVNCERLKALP  
TTKNWSIAMYFRFVIADYFIDQDKILYLDADIAQQGNLKLPTMDLANNVAAVVTET  
DANWNSLRAQSLQCNLELKGYNFSGVLLINTLAWAQESVSAKAMSLADKAVVSRLLTY  
MDQDILNLILGLKVFPIDAKYNTQFSLNYELKKSFCIPINDETVLIIHYVGPYKPHWYH  
AGYPSAQPFIKGKERASPWKNEPLMRPMNSNYARYCAKNFKQNKPINGIMNVIYFYFL  
KIHK"  
complement(3872..4678)  
/gene="rfap"  
/locus\_tag="c4454"  
complement(3872..4678)  
/gene="rfap"  
/locus\_tag="c4454"  
/function="enzyme; Macromolecule metabolism:  
Lipopolysaccharide"  
/note="Escherichia coli K-12 ortholog: b3630; Escherichia  
coli O157:H7 ortholog: z5054"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide core biosynthesis protein  
rfap"  
/protein\_id="AAN82890.1"  
/db\_xref="GI:26110706"  
/translation="MVMVELKEPFFATLWRGKDPFEVKTQGEVFRELETRTLRFE  
MAGKSYFLKWHRGTTLKEIIKNLLSLRMPVLGADREWNAIHELRLDVGVDVTWYGVAFGE  
KGINPLTSTFIITTEDLTPTISLEDCADWATNPDDVRVKRMLIKRVATMVRDMHAAG  
INHRDCYICHELLHLPFSGKEELKISVIDLHRAQLRTRVPRWRDKDLIGLYFSSMN  
IGLTORDIWRFMKVYFSAPLKDILKQEQGLLSQAEAKATKIRERTIRKSL"  
complement(4662..5786)  
/gene="rfag"  
/locus\_tag="c4455"  
complement(4662..5786)  
/gene="rfag"  
/locus\_tag="c4455"  
/function="enzyme; Macromolecule metabolism:  
Lipopolysaccharide"  
/note="Glucosyltransferase I; Escherichia coli K-12  
ortholog: b3631; Escherichia coli O157:H7 ortholog: z5055"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide core biosynthesis protein  
rfag"  
/protein\_id="AAN82891.1"  
/db\_xref="GI:26110707"  
/translation="MIVAFCLYKYFFPGGLQDFMRIAQTVAARGHHVRVVTQSWEGE  
CPDVFEELIKVPVKSHTNHGRNAYFAWQKHLREHPVDVRVGFNKMPLGLDVVYAADVC  
YAEKVAQEKGFYRLTSRYHYAAFERATFEQKPTQLMLTQDIADFKHYQTEAE  
RPHILPPIYVDRKYSQQPANSREIFRKNNGITEQQYLLQLVQSSDFTRKGVDRSIEAL  
ASLPSLRHNTLLYVVGQDKPKRFEALAEKRGVRSNVHFFSGRNDVSELMAAADLLH  
PAYQBAAGIVLLEAITAGLPVLTAVCGYAHYIVDANCGEAIAEPPFRQETLNEILRKA  
LTQSSLRQAWAENARHYADTDLYSLPEKAADIITGGLDG"  
complement(5783..6841)  
/gene="rfaq"  
/locus\_tag="c4456"  
complement(5783..6841)  
/gene="rfaq"

gene

CDS

AKRLNIPGMNGRYFNAGIYVNLKKWHEANLTPYLLTLRGETTKYGLSKYLDQDALNI

/locus\_tag="c4456"  
/EC\_number="2.-.-.-"  
/function="enzyme"; Macromolecule metabolism:  
Lipopolysaccharide  
/note="Escherichia coli K-12 ortholog: b3632; Escherichia coli O157:H7 ortholog: z5056"  
/codon\_start=1  
/transl\_table=11  
/product="Lipopolysaccharide core biosynthesis glycosyl transferase xfaq"  
/protein\_id="AAN82892.1"  
/db\_xref="GI:26110708"  
/translation="MDKFFRILLIKMRFHGMILLTTPVISSLKKNYPDAKIDVLLYQ DTIPILSENPEINALYGIKNKAKASEKIANFFYLIKVLRAKNDLIVNLTDQWVAI LRVLLNARVKISQDYHHRQSAFWRNSFTHLVPLQGGNVVESNLSVLPLGLESYKQT TMSYPPASWKRERELDHAGVQNYVVIQPTARQIFKWDNAKFSVIDALHARGYEV VLTSGPDKDDLACVNEIAQGCQTPPVTAAGKVTFFELGALIDHAQLFIGVDSAPAH IAAVNTPLISFGATDHIIFWRPWSNNMIQWAGDYREMPTRDQRDRNEMYSVIPAAD VIAAVDKLLPSSTGTSL"  
7254..8531  
/gene="kdtA"  
/locus\_tag="c4457"  
7254..8531  
/gene="kdtA"  
/locus\_tag="c4457"  
/EC\_number="2.-.-.-"  
/function="enzyme"; Surface polysaccharides and antigens"  
/note="Escherichia coli K-12 ortholog: b3633; Escherichia coli O157:H7 ortholog: z5057"  
/codon\_start=1  
/transl\_table=11  
/product="3-deoxy-D-manno-octulosonic-acid transferase"  
Query Match 92.1%; Score 566.4; DB 1; Length 301566;  
Best Local Similarity 96.9%; Pred. No. 1.9e-169;  
Matches 599; Conservative 0; Mismatches 16; Indels 3; Gaps 2;  
  
QY 1 ATGACCTTAGAATGGTGGTTGCCTACTGCTGACATCGATCATTTAAACGCTTCGCCA 60  
Db 281194 ATGACCTTAGAATGGTGGTTGCCTACTGCTGACATCGATCATTTAAACGCTTCGCCA 281135  
  
QY 61 GGCTCTGGTGGCAATCAACACTATGACCACTCGCTCACCACCGGTTAT--CCGSCCGGTG 118  
Db 281134 GGCTCTGGCGCAATCAACACATGACCACTCGCTCACCACCGGTTATCGCGGCGGTG 281075  
  
QY 119 GGGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTCGTGGTTGGCGTGGG 178  
Db 281074 GGGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTCGTGGTTGGCGTGGG 281015  
  
QY 179 TT-GGGACGCTATTTTCCCGTCAGTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCG 237  
Db 281014 TTGGGACGCTATTTTCCCGTCAGTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCGCA 280955  
  
QY 238 GCTTACTTGATTGGCTGGAATCCAGCAGTGGCGCGCTGGTGCATTTGACCTTAA 297  
Db 280954 GCTTACTTGATTGGCTGGAATACAGCAATGGCGCGCTGGCGCAATTGATCTTAA 280895  
  
QY 298 TCGCTGGCCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 357  
Db 280894 TCGCTGGCCTCCACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 280835  
  
QY 358 ACCAATCCAAAAGTATTGTGTTTCTGGCGCGCTATTCCGCAATTCATCATCGCGCA 417  
Db 280834 ACCAATCCAAAAGTATTGTGTTTCTGGCGCGCTATTCCGCAATTCATCATCGCGCAG 280775  
  
QY 418 CAGCCGCACTGATGCAGTATATCGTCTGGCGTACCACCTATTGTCGTGATATTATT 477  
Db 280774 CAGCCACAACTGATGCAGTATATCGTCTGGCGTACCACCTATTGTCGTGATATTATT 280715  
  
QY 478 GTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGACCAAG 537  
Db 280714 GTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGACCAAG 280655

QY 538 CAGATGAAGCGCGTGAATAAGATTTTCGGCTCGTTTGTATGCTGGTGGAGCGCTGTTA 597  
Db 280654 CAGATGAAGCGCGTGAATAAGATTTTCGGCTCGTTTGTATGCTGGTGGAGCGCTGTTA 280595  
  
QY 598 GCATCGGCGAGGCATGCG 615  
Db 280594 GCATCGGCGAGGCATGCG 280577  
  
RESULT 11  
AE005613/c  
LOCUS AE005613 10592 bp DNA linear BCT 21-MAR-2001  
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 232 of 290.  
ACCESSION AE005613 AE005174  
VERSION AE005613.1 GI:12518685  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli O157:H7 EDL933  
Escherichia coli O157:H7 EDL933  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 10592)  
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7  
JOURNAL Nature 409 (6819), 529-533 (2001)  
MEDLINE 21074935  
PUBMED 11206551  
REFERENCE 2 (bases 1 to 10592)  
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
TITLE Direct Submission  
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source  
1..10592  
/organism="Escherichia coli O157:H7 EDL933"  
/mol\_type="genomic DNA"  
/strain="EDL933"  
/serotype="O157:H7"  
/db\_xref="taxon:155864"  
/note="enterohemorrhagic"  
misc\_feature  
1..71  
/note="O-island #155; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"  
complement(72..836)  
/gene="Z5332"  
complement(72..836)  
/gene="Z5332"  
/function="orf; Unknown function"  
/note="Residues 1 to 131 of 254 are 95.41 pct identical to residues 1 to 131 of 161 from Escherichia coli K-12 Strain MG1655: B3815"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAG59008.1"  
/db\_xref="GI:12518686"  
/translation="MAHRLLIKGMITLNLKRIFLALTLLPLFAVAADDCALSDPTLT VQAYTVNPQTERVKMYQKANGAWGTLHALLADMSQGVQVMAMNGGIYDESAYPLG LYIENGQQKVALNLASGEGNFFIRPGGVFYAGDKVGIIVLDAFKTSKEIQFAVQSGP MLMNSVINPRIHPNVASRKIRNGVINKGNVAVLLSQATNFYDFACYAKAKLNE QLLYLDGTISHMYMKGGAIPWQRYPFVTMISVERKG"  
1206..2156  
/gene="corA"  
/note="synonym: Z5333"  
gene



CDS	1206. .2156 /gene="corA" /function="transport; Transport of small molecules; Cations" /note="Residues 1 to 316 of 316 are 100.00 pct identical to residues 1 to 316 of 316 from Escherichia coli K-12 Strain MG1655; B3816" /codon_start=1 /transl_table=11 /product="Mg2+ transport, system I" /protein_id="AAG59009.1" /db_xref="GI:12518687" /translation="MLSAFQLENNRLTRLEVEESQPLVNAVWIDLVEPDDDERLURVQS ELQSLATRPELEDIEASARFFEDDDGLHIHSFFFFEDAEHDAGNSTVAFTIRDGRLLF TLRERELPAFLRYMRARSQSMVDGNAYELLDDLFEFKIEQLADEENIYSDLEQLSR VMEGHQGYDEYDSEALSTLAELDIDGKVRICLMDTQALNFLVRKARLPGGQLEQARE ILRDIESLLPHNESLFQKVNFLMQAAMGFINIEQNRIRIKIFSUVSVVFLPPTLVASSY GNNFEFMPELKWSFGYPGAIIFMILAGLAPYLYFKRKNWL" 2157. .4848 /note="O-island #156; Region of the BDL933 chromosome not homologous to E. coli K-12 MG1655" complement(2198. .2689) /gene="Z5334" complement(2198. .2689) /gene="Z5334" /function="orf; Unknown function" /note="No significant matches" /codon_start=1 /transl_table=11 /protein_id="AAG59010.1" /db_xref="GI:12518688" /translation="MKSKLOKILLCFLCCIYNLWLRPVQILYTSYDAGNSVFLVW DHPWTSDSKINWYLKHQNEIKNOHPLPEGSWHTWYVIDIGNGFTDYKKYIEGPEYDL YCFPTIKSNDNCITKNYLMVINEYPYRNTHIGINDFTEYQLTQENKIEQVFNPHNFKK DNF" complement(2686. .3534) /gene="Z5335" complement(2686. .3534) /gene="Z5335" /function="orf; Unknown function" /note="No significant matches" /codon_start=1 /transl_table=11 /protein_id="AAG59011.1" /db_xref="GI:12518689" /translation="MNNTLPSINIPCTLFETISLFDNYSADDMQYGDWVEQDFLSLG LSDISAKVDYRLIKYHFPFGPGSINAFPSASSGKISQRECTDILFAEMKELAKMFS FFGQYKTLIEDLIEHFRYGNNGSNFHSQNLNLSFHEKINKYGNSPIRIIEKCIENGIN STPSTGYQPIQLQSIKTKLLSSRLNKENDFEDSFNGLGISVHDISAQKISLSFQRYA IGWSATIHFAQDHFGLDVTDIKNKTYSKYRFFRIWFFLQHRKDFAFKPFPTNENTIE RIENYL" 3748. .3867 /gene="Z5336" 3748. .3867 /gene="Z5336" /function="orf; Unknown function" /note="Residues 1 to 39 of 39 are 100.00 pct identical to residues 278 to 316 of 316 from GenPept 118 : gi 6960250 gb AAF33440.1  (AF233324) S. typhimurium magnesium transport protein (COR) (SP:P311138) [Salmonella typhimurium LT2]" /codon_start=1 /transl_table=11 /product="partial repeat of corA" /protein_id="AAG59012.1" /db_xref="GI:12518690" /translation="MNFEFMPELKWSFGYPGAIIFMILAGLAPYLYFKRKNWL" complement(3892. .4278) /gene="Z5337" complement(3892. .4278) /gene="Z5337" /function="orf; Unknown function" /note="No significant matches"	gene CDS
misc_feature	1206. .2156 /codon_start=1 /transl_table=11 /protein_id="AAG59013.1" /db_xref="GI:12518691" /translation="MFKSSIRINIFYPSKLERLSRFQFIYITLTWSILITMDINLAL SFILVINNTYVININIGAKRIDTGCKAIYFVLISLGLTLFSLISAIEGVRCELSLVQ LNVAWILFTTILMSAIPSAKMSLSAH" 4483. .4794 /gene="Z5339" 4483. .4794 /gene="Z5339" /function="orf; Unknown function" /note="No significant matches" /codon_start=1 /transl_table=11 /protein_id="AAG59014.1" /db_xref="GI:12518692" /translation="MDSIWEMVTFGSCSWVKNPYSKSDFKAWLLWNAITILVLFIL SLTTTLNRSIGIGLSFGFCQILGRYTMSLGWSKYWGVLGGFPVTNIVFLILIFFKK R" complement(4841. .5728) /gene="rarD" /note="synonym: Z5340" complement(4841. .5728) /gene="rarD" /function="orf; Unknown function" /note="Residues 4 to 295 of 295 are 92.12 pct identical to residues 8 to 299 of 300 from Escherichia coli K-12 Strain MG1655; B3819" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAG59015.1" /db_xref="GI:12518693" /translation="MDAKOTRQGVLLALAAFYIWIAGIAPAYEKLIIYVPADEILTHRVI WSFFFMVIMSIQRQWSYLKTLIQPKIFMLAVSAVILGIGNWLLFIWAVNNHHMLEA SLGYFINPLNIVLGMIFLGERFRMQWLAVILAICGVLQVLWTFGSLPIIALGLAFS FAFYGLVRKKIAVEAQTMLETMWLLPVAAYILFAIDSTSHMGQNPMSLNLLLLIA AGIVTTVPLLCFTAATRLRLSTLGFQYIGPTLMFLAVTFYGEKPGADKAVTFTFI WVALAIFVMDAIYTQRRKS" complement(5780. .6265) /gene="yigi" /note="synonym: Z5341" complement(5780. .6265) /gene="yigi" /function="orf; Unknown function" /note="Residues 1 to 161 of 161 are 100.00 pct identical to residues 1 to 161 of 161 from Escherichia coli K-12 Strain MG1655; B3820" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAG59016.1" /db_xref="GI:12518694" /translation="MKCKDDMSAVLTAEQALKLVGEMFYHMPFNRLGMELERYEKE FAQLAFKNQPMVGNWQAOSILHGGVIAASALDVAAGLVCVGSITLRHETISEDELRL SRWGTIDLRVDYLRPGRXERFTATSSLLRAGNKVAVARVELHNEEQLYIASATATYV G" 6412. .7281 /gene="plda" /note="synonym: Z5342" 6412. .7281 /gene="plda" /function="enzyme; Central intermediary metabolism: Phosphorus compounds"	gene CDS
gene	Query Match 90.0%; Score 553.8; DB 1; Length 10592; Best Local Similarity 97.2%; Pred. No. 1.2e-165; Matches 584; Conservative 1; Mismatches 13; Indels 3; Gaps 2;	QY 18 GTTTGCCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCAGGCTCTGGTGAATCAA 77       DB 10592 GTTTGCCTACCTACTGACATCAATCATTTTAAAGCCTGTCGCCAGGCTCTGGTGAATCAA 10533 

```
QY 78 CACTATGACCACCTCGCTCAACCACGGTTAT--CCGGCCGGTGGCGTCTATTGCTGGGCT 135
Db 10532 CACTATGACCACCTCGCTCAACCACGGTTATCGCGCGCGGTGGCGTCTATTGCGGGCT 10473
QY 136 TCAGACCGGACTGGCGATTTCATATTGCTGTTGGGTGGGTGGGT--GGGACGCTATTTC 194
Db 10472 TCAGACCGGACTGGCGATTTCATATTGCTGTTGGGTGGGTGGGTGGGACGCTATTTC 10413
QY 195 CCGCTCAGTGATTGCTTTGAAGTGTGAAGTGGGAGCGCGGCTTACTTGATTGGCT 254
Db 10412 CCGCTCAGTGATTGCTTTGAAGTGTGAAGTGGGAGCGCGGCTTACTTGATTGGCT 10353
QY 255 GGGAAATCCAGAGTGGCGCGCGCTGGTGCAATTGACCTTAAATCGCTGGCCTTACTCA 314
Db 10352 GGGAAATACAGCAATGGCGCGCGCTGGCGCAATTGATCTTAAATCGCTGGCCTTACTCA 10293
QY 315 ATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCAACAATCCCAAAAGTAT 374
Db 10292 ATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCAACAATCCCAAAAGTAT 10233
QY 375 TGTGTTTCTGGCGCGCTATTTCCGCAATTCATCGCCGCAACAGCGCAACTGATGCA 434
Db 10232 TGTGTTTCTGGCGCGCTATTTCCGCAATTCATCGCCGCAACAGCGCAACTGATGCA 10173
QY 435 GTATATCGTGCTGGCGGTCACCACTATTGTGTCGATATTATTGATGATCGGTTACGC 494
Db 10172 GTATATCGTGCTGGTGTCACCACTATTGTGTCGATATTATTGATGATCGGTTACGC 10113
QY 495 CACCCTTGCTCAACGGATTGCTCTCTATGGATTAAAGACCAAGAGATGAAGCGCTGAA 554
Db 10112 CACCCTTGCTCAACGGATTGCTCTCTATGGATTAAAGACCAAGAGATGAAGCGCTGAA 10053
QY 555 TAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTAGCATCGGCGAGGCATGC 614
Db 10052 TAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTAGCATCGGCGAGGCATGC 9993
QY 615 G 615
Db 9992 G 9992
```

```
RESULT 12
AC020970/c
LOCUS AC020970 212936 bp DNA linear HTG 10-FEB-2000
DEFINITION Mus musculus clone RP23-252M21, WORKING DRAFT SEQUENCE, 144
unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation
-----
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 144 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 1159: contig of 1159 bp in length
gap of unknown length
1160 2363: contig of 1204 bp in length
gap of unknown length
2364 3604: contig of 1241 bp in length
gap of unknown length
3605 5188: contig of 1584 bp in length
gap of unknown length
5189 6372: contig of 1184 bp in length
gap of unknown length
6373 7737: contig of 1365 bp in length
gap of unknown length
7738 9783: contig of 2046 bp in length
gap of unknown length
9784 11046: contig of 1263 bp in length
gap of unknown length
11047 12223: contig of 1177 bp in length
gap of unknown length
12224 13492: contig of 1269 bp in length
gap of unknown length
13493 14669: contig of 1177 bp in length
gap of unknown length
14670 15886: contig of 1217 bp in length
gap of unknown length
15887 17089: contig of 1203 bp in length
gap of unknown length
17090 18308: contig of 1219 bp in length
gap of unknown length
18309 19518: contig of 1210 bp in length
gap of unknown length
19519 20710: contig of 1192 bp in length
gap of unknown length
20711 21932: contig of 1222 bp in length
gap of unknown length
21933 23606: contig of 1674 bp in length
gap of unknown length
23607 25328: contig of 1722 bp in length
gap of unknown length
25329 26649: contig of 1321 bp in length
gap of unknown length
26650 27849: contig of 1200 bp in length
gap of unknown length
27850 29135: contig of 1286 bp in length
gap of unknown length
29136 30700: contig of 1565 bp in length
gap of unknown length
30701 31877: contig of 1177 bp in length
gap of unknown length
31878 33065: contig of 1188 bp in length
gap of unknown length
33066 34267: contig of 1202 bp in length
gap of unknown length
34268 35474: contig of 1207 bp in length
gap of unknown length
35475 36650: contig of 1176 bp in length
gap of unknown length
36651 37859: contig of 1209 bp in length
gap of unknown length
37860 39072: contig of 1213 bp in length
gap of unknown length
39073 40406: contig of 1334 bp in length
gap of unknown length
40407 41630: contig of 1224 bp in length
gap of unknown length
41631 42864: contig of 1234 bp in length
```

\* 42865 gap of unknown length  
\* 44200: contig of 1336 bp in length  
\* 44201 gap of unknown length  
\* 46087: contig of 1887 bp in length  
\* 46088 gap of unknown length  
\* 48227: contig of 2140 bp in length  
\* 48228 gap of unknown length  
\* 49518: contig of 1291 bp in length  
\* 49519 gap of unknown length  
\* 50910: contig of 1392 bp in length  
\* 50911 gap of unknown length  
\* 52420: contig of 1510 bp in length  
\* 52421 gap of unknown length  
\* 53613: contig of 1193 bp in length  
\* 53614 gap of unknown length  
\* 55004: contig of 1391 bp in length  
\* 55005 gap of unknown length  
\* 56300: contig of 1296 bp in length  
\* 56301 gap of unknown length  
\* 57505: contig of 1205 bp in length  
\* 57506 gap of unknown length  
\* 58686: contig of 1181 bp in length  
\* 60407: gap of unknown length  
\* 60408 contig of 1721 bp in length  
\* 61632: gap of unknown length  
\* 61633 gap of unknown length  
\* 62880: contig of 1248 bp in length  
\* 64220: gap of unknown length  
\* 64221 gap of unknown length  
\* 65420: contig of 1200 bp in length  
\* 65421 gap of unknown length  
\* 66605: contig of 1185 bp in length  
\* 66606 gap of unknown length  
\* 68682: contig of 2077 bp in length  
\* 68683 gap of unknown length  
\* 70712: contig of 2030 bp in length  
\* 71895: gap of unknown length  
\* 71896 contig of 1183 bp in length  
\* 73462: gap of unknown length  
\* 73463 contig of 1567 bp in length  
\* 74806: gap of unknown length  
\* 74807 contig of 1344 bp in length  
\* 76205: gap of unknown length  
\* 76206 contig of 1399 bp in length  
\* 77580: gap of unknown length  
\* 77581 contig of 1375 bp in length  
\* 79082: gap of unknown length  
\* 80314: contig of 1501 bp in length  
\* 80315 gap of unknown length  
\* 81899: gap of unknown length  
\* 81900 contig of 1375 bp in length  
\* 83140: gap of unknown length  
\* 83141 contig of 1501 bp in length  
\* 84665: gap of unknown length  
\* 84666 contig of 1241 bp in length  
\* 85888: gap of unknown length  
\* 85889 contig of 1525 bp in length  
\* 87155: gap of unknown length  
\* 87156 contig of 1223 bp in length  
\* 88435: gap of unknown length  
\* 88436 contig of 1267 bp in length  
\* 89955: gap of unknown length  
\* 89956 contig of 1520 bp in length  
\* 91277: gap of unknown length  
\* 91278 contig of 1322 bp in length  
\* 92513: gap of unknown length  
\* 92514 contig of 1236 bp in length  
\* 93917: gap of unknown length  
\* 93918 contig of 1404 bp in length  
\* 95448: gap of unknown length  
\* 95449 contig of 1336 bp in length  
\* 96800: gap of unknown length  
\* 96801 contig of 1887 bp in length  
\* 97978: gap of unknown length  
\* 97979 contig of 2140 bp in length  
\* 99179: gap of unknown length  
\* 99180 contig of 1291 bp in length  
\* 100437: gap of unknown length  
\* 100438 contig of 1392 bp in length  
\* 101828: gap of unknown length  
\* 101829 contig of 1510 bp in length  
\* 103117: gap of unknown length  
\* 103118 contig of 1193 bp in length  
\* 104550: gap of unknown length  
\* 104551 contig of 1391 bp in length  
\* 105854: gap of unknown length  
\* 105855 contig of 1296 bp in length  
\* 107505: gap of unknown length  
\* 107506 contig of 1205 bp in length  
\* 108686: gap of unknown length  
\* 111811: contig of 1181 bp in length  
\* 111812 gap of unknown length  
\* 111813 contig of 1721 bp in length  
\* 111814 gap of unknown length  
\* 111815 contig of 1225 bp in length  
\* 111816 gap of unknown length  
\* 111817 contig of 1248 bp in length  
\* 111818 gap of unknown length  
\* 111819 contig of 1340 bp in length  
\* 111820 gap of unknown length  
\* 111821 contig of 1200 bp in length  
\* 111822 gap of unknown length  
\* 111823 contig of 1185 bp in length  
\* 111824 gap of unknown length  
\* 111825 contig of 2077 bp in length  
\* 111826 gap of unknown length  
\* 111827 contig of 2030 bp in length  
\* 111828 gap of unknown length  
\* 111829 contig of 1183 bp in length  
\* 111830 gap of unknown length  
\* 111831 contig of 1567 bp in length  
\* 111832 gap of unknown length  
\* 111833 contig of 1344 bp in length  
\* 111834 gap of unknown length  
\* 111835 contig of 1399 bp in length  
\* 111836 gap of unknown length  
\* 111837 contig of 1375 bp in length  
\* 111838 gap of unknown length  
\* 111839 contig of 1501 bp in length  
\* 111840 gap of unknown length  
\* 111841 contig of 1375 bp in length  
\* 111842 gap of unknown length  
\* 111843 contig of 1501 bp in length  
\* 111844 gap of unknown length  
\* 111845 contig of 1241 bp in length  
\* 111846 gap of unknown length  
\* 111847 contig of 1525 bp in length  
\* 111848 gap of unknown length  
\* 111849 contig of 1223 bp in length  
\* 111850 gap of unknown length  
\* 111851 contig of 1267 bp in length  
\* 111852 gap of unknown length  
\* 111853 contig of 1280 bp in length  
\* 111854 gap of unknown length  
\* 111855 contig of 1520 bp in length  
\* 111856 gap of unknown length  
\* 111857 contig of 1322 bp in length  
\* 111858 gap of unknown length  
\* 111859 contig of 1236 bp in length  
\* 111860 gap of unknown length  
\* 111861 contig of 1404 bp in length  
\* 111862 gap of unknown length

\* 93918 95448: contig of 1531 bp in length  
\* 95449 gap of unknown length  
\* 96800: contig of 1352 bp in length  
\* 96801 gap of unknown length  
\* 97978: contig of 1178 bp in length  
\* 97979 gap of unknown length  
\* 99179: contig of 1201 bp in length  
\* 99180 gap of unknown length  
\* 100437: contig of 1258 bp in length  
\* 100438 gap of unknown length  
\* 101827: contig of 1390 bp in length  
\* 101828 gap of unknown length  
\* 103117: contig of 1290 bp in length  
\* 103118 gap of unknown length  
\* 104550: contig of 1433 bp in length  
\* 104551 gap of unknown length  
\* 105854: contig of 1304 bp in length  
\* 105855 gap of unknown length  
\* 107523: contig of 1669 bp in length  
\* 107524 gap of unknown length  
\* 109208: contig of 1685 bp in length  
\* 109209 gap of unknown length  
\* 110430: contig of 1222 bp in length  
\* 110431 gap of unknown length  
\* 111662: contig of 1232 bp in length

Query Match 75.9%; Score 467; DB 2; Length 212936;  
Best Local Similarity 96.4%; Pred. No. 1.3e-137;  
Matches 595; Conservative 0; Mismatches 10; Indels 12; Gaps 11;

QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCAATTTTAAACGCTGTGCGCA 60  
Db 175539 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTG-CATCGATCAATTTTAAAGCCTGTGCGCA 175481  
QY 61 GGCTCTGGTGAATCAACACTATGACCACTGCTGCTCAACCCAGGTTAT--CCGGCCGGTG 118  
Db 175480 GGCTCTGGTGAATCAACACTAT-ACCACCTGCTCAACCCAGGTTATCGCGGCGGGTG 175422  
QY 119 GCGTCTATTGCTGGGTTTCAGACCGGACTGGCGATTCAATTTGCTGCTGGTGGGTTGGG 178  
Db 175421 GCGTCTATTGCT-GGCTTCAGACCGGACTGGCGATTCAATTTGCTGCTGGTGGGTTGGG 175363  
QY 179 TTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGAGGCGCGG 238  
Db 175362 TTGGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGAGGCGG 175304  
QY 239 CTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTCAATTTGACCTTAAAT 298  
Db 175303 CTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTCAATTTGACCTTAAAT 175244  
QY 299 CGCTGGCTCTACTCAATCGCTCGACATTTTGTCCAGCGCGAGTTTGTGAAATCTCA 358  
Db 175243 CGCTGGCTCTACTCAATCGCTCGACATTT-TTCCAGCGCGAGTTTGTGAAATCTCA 175185  
QY 359 CCAATCCCAAAAGTATTGTGTTTCTGGCGGCGGTATTTCCGCAATTCATCATGCCGCAAC 418  
Db 175184 CCAATCCCAAAAGTATTGTG-TTCTGGCGGCGGTATTTCCGCAATTCATCATGCCGCAAC 175126  
QY 419 AGCCGCAACTGATGAGTATATCCTGCTCGGCTCACCACCTATTGTGGTCAATATTG 478  
Db 175125 AGCCGCAACTG-TGCAGTATATCCTGCTCGGCTCACCACCTATTGTGGTCAATATTG 175067  
QY 479 TGATGATCGGTTACGCCACCTTCTCTCAACCGGATTGCTCTATGGATTAAAGCAACAAAGC 538  
Db 175066 T-ATGATCGGTTACGCCACCTTCTCTCAACCGGATTGCTCTATGGATTAAAGCAACAAAGC 175009  
QY 539 AGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGGTGGGAGCGCTGTAG 598  
Db 175008 AGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATG-TGGTGGGAGCGCTGTAG 174950  
QY 599 CATCGGCGAGGCATGCG 615  
Db 174949 CATCGGCGAGGCATGCG 174933

RESULT 13  
AC020870/c  
LOCUS AC020870 256373 bp DNA linear HTG 17-FEB-2000  
DEFINITION Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC020870  
VERSION AC020870.2 GI:6984372  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 256373)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 256373)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 17, 2000 this sequence version replaced gi:6686438.  
\* NOTE: This record contains 198 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1 752: contig of 752 bp in length  
\* gap of unknown length  
\* 753 1000: contig of 248 bp in length  
\* gap of unknown length  
\* 1001 1604: contig of 604 bp in length  
\* gap of unknown length  
\* 1605 2302: contig of 698 bp in length  
\* gap of unknown length  
\* 2303 2699: contig of 397 bp in length  
\* gap of unknown length  
\* 2700 3686: contig of 987 bp in length  
\* gap of unknown length  
\* 3687 4099: contig of 413 bp in length  
\* gap of unknown length  
\* 4100 4778: contig of 679 bp in length  
\* gap of unknown length  
\* 4779 5571: contig of 793 bp in length  
\* gap of unknown length  
\* 5572 5967: contig of 396 bp in length  
\* gap of unknown length  
\* 5968 6763: contig of 796 bp in length  
\* gap of unknown length  
\* 6764 7639: contig of 876 bp in length  
\* gap of unknown length  
\* 7640 8442: contig of 803 bp in length  
\* gap of unknown length  
\* 8443 9414: contig of 972 bp in length  
\* gap of unknown length  
\* 9415 10236: contig of 822 bp in length  
\* gap of unknown length  
\* 10237 10844: contig of 608 bp in length  
\* gap of unknown length  
\* 10845 11504: contig of 660 bp in length  
\* gap of unknown length  
\* 11505 12321: contig of 817 bp in length  
\* gap of unknown length  
\* 12322 12652: contig of 331 bp in length  
\* gap of unknown length  
\* 12653 14116: contig of 1464 bp in length

14117 gap of unknown length  
15298: contig of 1182 bp in length  
15299 gap of unknown length  
16234: contig of 936 bp in length  
16235 gap of unknown length  
16376: contig of 142 bp in length  
16664: contig of 288 bp in length  
17631: contig of 967 bp in length  
17632 gap of unknown length  
19241: contig of 1610 bp in length  
20453: contig of 1212 bp in length  
20559: contig of 106 bp in length  
21117: contig of unknown length  
21655: contig of 538 bp in length  
21818: contig of 163 bp in length  
22679: contig of 861 bp in length  
23579: contig of 900 bp in length  
24383: contig of 804 bp in length  
25612: contig of 1229 bp in length  
25957: contig of 345 bp in length  
27260: contig of 1303 bp in length  
28596: contig of 1336 bp in length  
29473: contig of 877 bp in length  
30372: contig of 899 bp in length  
31202: contig of 830 bp in length  
32688: contig of 1486 bp in length  
33431: contig of 743 bp in length  
33693: contig of 262 bp in length  
34616: contig of 923 bp in length  
35884: contig of 1268 bp in length  
36459: contig of 575 bp in length  
36996: contig of 537 bp in length  
37941: contig of 945 bp in length  
38682: contig of 741 bp in length  
39549: contig of 867 bp in length  
40792: contig of 1243 bp in length  
42585: contig of 1793 bp in length  
44115: contig of 1530 bp in length  
45300: contig of 1185 bp in length  
46417: contig of 1117 bp in length  
gap of unknown length

46418 47815: contig of 1398 bp in length  
\* gap of unknown length  
\* 47816 49310: contig of 1495 bp in length  
\* gap of unknown length  
\* 49311 50304: contig of 994 bp in length  
\* gap of unknown length  
\* 50305 50848: contig of 544 bp in length  
\* gap of unknown length  
\* 50849 51828: contig of 980 bp in length  
\* gap of unknown length  
\* 51829 52699: contig of 871 bp in length  
\* gap of unknown length  
\* 52700 53365: contig of 666 bp in length  
\* gap of unknown length  
\* 53366 54633: contig of 1268 bp in length  
\* gap of unknown length  
\* 54634 55470: contig of 837 bp in length  
\* gap of unknown length  
\* 55471 56358: contig of 888 bp in length  
\* gap of unknown length  
\* 56359 56934: contig of 576 bp in length  
\* gap of unknown length  
\* 56935 57672: contig of 738 bp in length  
\* gap of unknown length  
\* 57673 58319: contig of 647 bp in length  
\* gap of unknown length  
\* 58320 58529: contig of 210 bp in length  
\* gap of unknown length  
\* 58530 59083: contig of 554 bp in length  
\* gap of unknown length  
\* 59084 59906: contig of 823 bp in length  
\* gap of unknown length  
\* 59907 60826: contig of 920 bp in length  
\* gap of unknown length  
\* 60827 61084: contig of 258 bp in length  
\* gap of unknown length  
\* 61085 61762: contig of 678 bp in length  
\* gap of unknown length  
\* 61763 62638: contig of 876 bp in length  
\* gap of unknown length  
\* 62639 62713: contig of 75 bp in length  
\* gap of unknown length  
\* 62714 63593: contig of 880 bp in length  
\* gap of unknown length  
\* 63594 64039: contig of 446 bp in length  
\* gap of unknown length  
\* 64040 65076: contig of 1037 bp in length  
\* gap of unknown length  
\* 65077 65366: contig of 290 bp in length  
\* gap of unknown length  
\* 65367 65590: contig of 224 bp in length  
\* gap of unknown length  
\* 65591 66041: contig of 451 bp in length  
\* gap of unknown length  
\* 66042 67040: contig of 999 bp in length  
\* gap of unknown length  
\* 67041 67560: contig of 520 bp in length  
\* gap of unknown length

Query Match 73.1%; Score 449.8; DB 2; Length 256373;  
Best Local Similarity 95.7%; Pred. No. 4.5e-132;  
Matches 495; Conservative 0; Mismatches 17; Indels 5; Gaps 3;  
QY 1 ATGACCTTAGAATGGTGGTTGCCCTACCTGCTGACATCGATCAATTTAA--CGGTGTCGC 58  
DB 77559 ATGACCTTAGAATGGTGGTTGCCCTACCTGCTGACATCGATCAATTTAAGCCTCGTCGCC 77500  
QY 59 CAGGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACACCGGTTAT--CCGGCCGG 116  
DB 77499 CAGGCTCTGGTGCAATCAACACTATGACCACCTCGCTCAACACCGGTTATCGCGCGCG 77440  
QY 117 TGGCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTCAATTTGCTGCTGGTGGCGTGG 176

Db 77439 TGGCGTCTATGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTGCTGTTGGCGTGG 77380  
QY 177 GGTG-GGGACGCTATTTTCCCGCTCAGTGATTCGTTTGAAGTGTGGAAGTGGCAGGCG 235  
Db 77379 GGTGCGGACGCTATTTTCCCGCTCAGTGATTCGTTTGAAGTGTGGAAGTGGCAGGCG 77320  
QY 236 CGGCTTACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGAATTCACCTTA 295  
Db 77319 CGGCTTACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGAATTCACCTTA 77260  
QY 296 AATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATC 355  
Db 77259 AATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATC 77200  
QY 356 TCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGC 415  
Db 77199 TCACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGC 77140  
QY 416 AACAGCGCGCACTGATGCAGTATATCGTGTCTGGCGCTCACCACCTATTGTTGGTCGATATA 475  
Db 77139 AACAGCGCGCACTGATGCAGTATATCGTGTCTGGCGCTCACCACCTATTGTTGGTCGATATA 77080  
QY 476 TTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGAT 512  
Db 77079 TTGTGATGATCGGTTACGCCACCCCTTGCTCAACGGAT 77043

RESULT 14  
AR352333 LOCUS AR352333 417 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 274 from patent US 6589738.  
ACCESSION AR352333  
VERSION AR352333.1 GI:33757296  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 417)  
AUTHORS Forsyth,R.A., Ohlsen,K. and Zyskind,J.W.  
TITLE Genes essential for microbial proliferation and antisense thereto  
JOURNAL Patent: US 6589738-A 274 08-JUL-2003;  
FEATURES  
Source 1. 417  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 67.3%; Score 414; DB 6; Length 417;  
Best Local Similarity 100.0%; Pred. No. 4.6e-121;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 202 GTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCGGCTTACTTGTGTTGGCTGGGAATC 261  
Db 1 GTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCGGCTTACTTGTGTTGGCTGGGAATC 60  
QY 262 CAGCAGTGGCGCGCGCTGTTGCAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT 321  
Db 61 CAGCAGTGGCGCGCGCTGTTGCAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT 120  
QY 322 CGACATTTGTTCCAGCGCGCAGTGTGTTGTAATCTCACCACCTCCCAAAAGTATTGTGTTT 381  
Db 121 CGACATTTGTTCCAGCGCGCAGTGTGTTGTAATCTCACCACCTCCCAAAAGTATTGTGTTT 180  
QY 382 CTGGCGCGCGCTATTTCGCAATTTCATCATGCGCAACAGCGCGCAACTGATGCAGTATATC 441  
Db 181 CTGGCGCGCGCTATTTCGCAATTTCATCATGCGCAACAGCGCGCAACTGATGCAGTATATC 240  
QY 442 GTGCTCGGCGTCAACACTATTGTGGTCGATATTAATTGATGATCGGTTACGCCACCCCTT 501  
Db 241 GTGCTCGGCGTCAACACTATTGTGGTCGATATTAATTGATGATCGGTTACGCCACCCCTT 300  
QY 502 GCTCAACGGATTGCTCTATGGATTAAAGACCAAGCAGATGAAGCGCTGAATAGATT 561



Db 301 GCTCAACGGATTGCTCTATGATTAAAGGACCAAGCAGATGAAGGCGCTGAATAAGATT 360

QY 562 TTCCGCTCGTTGTTTATGCTGCTGGAGCGCTGTTAGCATCGGCGAGCGCATGCG 615  
|||||

Db 361 TTCCGCTCGTTGTTTATGCTGCTGGAGCGCTGTTAGCATCGGCGAGCGCATGCG 414  
|||||

RESULT 15  
AL627278 258050 bp DNA linear BCT 04-JUL-2003  
LOCUS  
DEFINITION  
Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,  
complete chromosome; segment 14/20.  
ACCESSION  
AL627278 AL513382  
VERSION  
AL627278.1 GI:16504263  
KEYWORDS  
SOURCE  
ORGANISM  
Salmonella enterica subsp. enterica serovar Typhi  
Salmonella enterica subsp. enterica serovar Typhi  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
REFERENCE  
1 (bases 1 to 258050)  
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,  
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,  
Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,  
Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,  
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,  
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,  
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,  
Stevens,K., Whitehead,S. and Barrell,B.G.  
Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18  
Nature 413 (6858), 848-852 (2001)  
21534947  
11677608  
2 (bases 1 to 258050)  
Parkhill,J.  
Direct Submission  
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella  
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK  
E-mail: parkhill@sanger.ac.uk  
Notes:  
Details of S. typhi sequencing at the Sanger Centre are available  
on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).  
Location/Qualifiers  
1. 258050  
/organism="Salmonella enterica subsp. enterica serovar  
Typhi"  
/mol type="genomic DNA"  
/strain="CT18"  
/db xref="taxon:90370"  
69..1544  
/gene="tolC"  
/note="synonym: STY3364"  
69..1544  
/gene="tolC"  
/note="Similar to Salmonella enteritidis outer membrane  
protein TolC precursor tolC SW:TOLC\_SALEN (Q54001) (491  
aa) fasta scores: E(): 0, 99.0% id in 491 aa, and to  
Escherichia coli outer membrane protein tolC precursor  
tolC or mtcb or muka or ref1 SW:TOLC\_ECOLI (P02930) (495  
aa) fasta scores: E(): 0, 89.7% id in 495 aa  
Orthologue of E. coli tolC (TOLC\_ECOLI); Fasta hit to  
TOLC\_ECOLI (495 aa), 90% identity in 495 aa overlap"  
/codon start=1  
/transl table=11  
/product="outer membrane protein TolC precursor"  
/protein id="CAD07712.1"  
/db\_xref="GI:16504264"  
/db\_xref="GOA:Q8Z3N8"  
/db\_xref="SPTRMBL:Q8Z3N8"  
/translation="WQMKLLPILIGLSLGSFSTLSQAEINLMQVYQQARLSNPRLKRS  
AADRDAAFEKINEARSPLLPOLGLGADYTYNSGYRDANGINSNETSASLQTLTFLDM

SKWRGLTLQEKAGIQDVTYQTDOOTLILNTANAYFKVLNAIDVLSYTOAQKEAIYRQ  
LDQTTORFNVGLVAITDVQNARAQYDTVLANEVNTARNLNDNAVELRQVITGNYYPELA  
SLNVEHFKTDKPKAVNALLKEAENRNLSLQARLSQDLAREQIRQAQDGHLPILNLTA  
STGISDTSYSGSKTNSAQYDDSNMGQNKIGLNFSLPLYQGGMVNSQVKQAQNFVGAS  
EQLGSAHRSVVQTVRSSEFNINASSINAYKQAVVSAQSSILDAMEAGYSVGTRTIVD  
VLDAITTLYDAKQQLANARYTYLINQLNLIKYALGTLINEQDLLALNSTLGKPIPTSPES  
VAPETPEQDAADGYNAHSAAPAVQPTAARANSNNGNPFRRH"  
1757..2428  
/gene="STY3365"  
1757..2428  
/gene="STY3365"  
/note="Orthologue of E. coli ygiB (YGIB\_ECOLI); Fasta hit  
to YGIB\_ECOLI (234 aa), 97% identity in 223 aa overlap.  
Contains a possible N-terminal signal sequence."  
/codon start=1  
/transl table=11  
/product="possible lipoprotein"  
/protein id="CAD07713.1"  
/db\_xref="GI:16504265"  
/db\_xref="SPTRMBL:Q8XGZ1"  
/translation="MKRTKSIHHASFRKWSARHLTPVALAVTAVFMLAGCEKSDSTV  
SLYQNADDCSAANPGKSAECTTAYNNALKEARTAPKYATREDCVAEFGEQCOQAPA  
QAGMAPENQAQAQSSGSFWMPLMAGYMMGLMGGAGFAQQPLFSSKNPASPAYGKY  
TDAAGKNYGAAPQGRMTVPKTAMAPKPATTTTVTRGGFGESVAKQSTMORSAAGTST  
RSMGG"  
1835..1867  
/gene="STY3365"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid  
attachment site"  
2434..3597  
/gene="STY3366"  
2434..3597  
/gene="STY3366"  
/note="Fasta hit to YJFC\_ECOLI (387 aa), 50% identity in  
392 aa overlap  
Orthologue of E. coli ygiC (YGIC\_ECOLI); Fasta hit to  
YGIC\_ECOLI (386 aa), 94% identity in 386 aa overlap"  
/codon start=1  
/transl table=11  
/product="conserved hypothetical protein"  
/protein id="CAD07714.1"  
/db\_xref="GI:16504266"  
/db\_xref="SPTRMBL:Q8XFP6"  
/translation="NERVSIATERDWRDKATEYGFNFHTMYGEPYWCEDAYKLTLAQ  
VEKLEDVTAELHQMKLVVERVIADELMTKFRIPKHTQVFRQSQWQTPQSPSLYSRLD  
LAWDGICEPKLLENNADTPTSLYEAFFQWILEQINAGNLPESDQFNLSQEKLIE  
RFAELREQYGFQLHLTCRDTVEDRGTIQYLDQCAAEIAIEFLYIDDIGEKGQ  
FTDLQDVIANLFKLYPWEFMLEMREMFSTKLEDAGVRWLEPAWKSII SNKALLPLLWEM  
FPDHPNLLPAYFAEDEHPMDKYVYKPIFSREGANVSIIENGKTIESVEGPGYEEGMI  
VQFYPLPKFGDSYTLIGSWLINDQPAIGIGIRDRALITQDLSRFYPHFIVEG"  
complement(3659..4489)  
/gene="STY3367"  
complement(3659..4489)  
/gene="STY3367"  
/note="Orthologue of E. coli ygiD (YGID\_ECOLI); Fasta hit  
to YGID\_ECOLI (271 aa), 87% identity in 266 aa overlap"  
/codon start=1  
/transl table=11  
/product="conserved hypothetical protein"  
/protein id="CAD07715.1"  
/db\_xref="GI:16504267"  
/db\_xref="GOA:Q8Z3N7"  
/db\_xref="SPTRMBL:Q8Z3N7"  
/translation="MWMLTKCLCKDKNIMSLTOMPALFLGHGSPMNVLDNDYTRAWR  
RLGEALPRPQAIIVVVSAAHWYTRGTGTAMERPOTLHDFGFPQALYDMHYPRPGSPAL  
AQLVELLAPVPVALDKEAWGFDHSGWGLIKMYPNADI PMVQLSVDSSTKPAWHFEM  
GRKLATLRDEGVMLVASGNVNHNLRTVRWHGDNIPYPWAASFNDVFKANLWQGPVEQ  
HPLVNYLQHEGGALSNTPEHFLPLLYVLGAWDGKEPITIPVDGIEMGSISMLSVQVG  
"  
4588..5361  
/gene="STY3368"  
4588..5361  
/gene="STY3368"  
gene  
CDS

/note="Orthologue of E. coli ygiE (YGIE\_ECOLI); Fasta hit to YGIE\_ECOLI (257 aa), 93% identity in 257 aa overlap. Contains multiple possible membrane spanning hydrophobic domains"

/codon\_start=1

/transl\_table=11

/product="putative membrane protein"

/protein\_id="CAD07716.1"

/db\_xref="GI:16504268"

/db\_xref="GOA:Q8XGR4"

/db\_xref="SWISS-PROT:Q8XGR4"

/translation="MSVPLILTLLAGAATFICAFGLVGLQKPSNRVLAFSLGFAAGIMLLISLMEMLPAAALDTGEMSPVLGYGMFIIIGLLGYGLDBLLPHAHPODLVQKROQLPFGSIKRTALLTLGLSHNFPPEGIATFTASNLGLFGALALAHNIPEGLAVAGPVYAATGSKRTAIFWAGTSGMAEILGGVLAWLILGSLVSPVMAAIIAAVAGIMVALSVDELMPLAKEIDPNNPNSYGLVCGMSIMGLSLVILQTIGIG"

6004.7800

/gene="STY3370"

/note="synonym: asst"

6004.7800

/gene="STY3370"

/note="Similar to Klebsiella pneumoniae arylsulfate sulfotransferase asst TR:P97036 (EMBL:U32616) (598 aa) fasta scores: E(): 0, 86.8% id in 598 aa"

/codon\_start=1

/transl\_table=11

/product="probable arylsulfate sulfotransferase"

/protein\_id="CAD07717.1"

/db\_xref="GI:16504269"

/db\_xref="GOA:Q8Z3N6"

/db\_xref="SPTRMBL:Q8Z3N6"

/translation="MFDDQYKRTILAGAVALTCLTAASTFAAGFOPAQAQPAQKLGAVVDPYGNAPLALTALVELDSHIIISDKVTVHGKGEKGVPTVTVGKESLEYDGIPIFGLYKFNANNVVEYKENGKAMKDDYVQTSAINVHMYMNRSSDLQQTQVIVKAPGFEDRLYLVNTHFTPQGAEEFHWGKDKNAGILDAGPAGGALPDIAPTYTFVVDQTQGEYRWLLDQDTFYNGHDMINIKRGYLMGIRETPRGFTTAVQGWYEFDMNGQIADHLKPRGLFASHTSETVNGTIVLLRVGKRDYRKEDGIHVHTIRDQIIEVDKSGRVVDWDLTKILD MRDALGALDAGAVCNVDLAHAGQQAULEPDTPTYGALGVGAGRNWAHVNSIAYDAKDDSIILSSRHQIVKIGRDKQVKWILAPSKGNKQLASKLLKPVDDHGKPLTCDENGKCKTDFDFTYQHTAWLSKGLTVFDNGDGRGLEQQLPMTKYSRPFVEYKIDKKGTVQVQWEYKGERGYDFSPITSVVEYQKDRDTMFQGGSGINLFDVGKPTVQKLEINIDYKTKEVKEIDVLSDXPNQTHYRALLVHPTQMKX"

7820.8491

/gene="dsba"

/note="synonym: STY3371"

7820.8491

/gene="dsba"

/note="Similar to Enterobacter amnigenus disulfide isomerase dsba TR:Q9XDP1 (EMBL:AF012826) (222 aa) fasta scores: E(): 0, 90.1% id in 222 aa"

/codon\_start=1

/transl\_table=11

/product="disulfide isomerase"

/protein\_id="CAD07718.1"

/db\_xref="GI:16504270"

/db\_xref="GOA:Q8Z3N5"

/db\_xref="SPTRMBL:Q8Z3N5"

/translation="MSSKWTSLFSKVVLTAAALVTPFAASAFTEGDTYMWLEKPIPNADKTLTKVFSYACPCYKYDKAVTGPVSDKVADLVTFPFHLETXGEYKQASVEFAVLIAKDKAAGISLFDPAKSQFKKAKFAWYTAYHDKKERSDQKPAPIKTGLDAGMSQADFEALKDPAVQETLEKWAAYDVAKIQGVPAVYVNGKYLIIYTKNKSIDSMAELVRELATKX"

7847.8488

misc feature

Db	247804	GGTTCAGCGCCATCAATACCATGACGACGTCTATCAACCATGGATATCGTGGCGCAGCG	247863
QY	119	GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTGCTGCTGGTGGCGTGGGG	178
Db	247864	GCTTCTATCGCCGGCTCCAGACCGGGCTGGGATACATATCGTACTGTTGGGCGTCCGA	247923
QY	179	TTGGG-ACGCTATTTTCCCGCTCAGTGAATTGCGTTTGAAGTGTGAAGTGGGCAGGCGCG	237
Db	247924	CTGGGTACGCTCTTTTCGGCTCGCTCATCGCTTTTGAATCTGAAATGGGCTGGCGCG	247983
QY	238	GCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAA	297
Db	247984	GCTTATCTATCTGGCTGGGTATCCAGCAATGGCGCGCAGGCGCTATCGATCTGCAT	248043
QY	298	TCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC	357
Db	248044	ACTCTGCCCAGACGCAATCGCGGGTCGGCTGTTTCAACGGGCGATATTTGTCAATCTA	248103
QY	358	ACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATGCCGCA	417
Db	248104	ACCAATCCCAAAAGTATTGTCTTTCTTGGCGCCCTGTTTCCGCAATTCATCATGCCGCG	248163
QY	418	CAGCCGCAACTGATGCAGTATATCGTCTCGGCGTCACCACTATTGTGTCGATATTATT	477
Db	248164	CAACCGCAACTGGCGCAGTACCTCATCTCGGCGTCACCAAGATTGTTGGTGGATATGATT	248223
QY	478	GTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG	537
Db	248224	GTGATGACCGGTTACGCCACCGCTGGCGCAGCGCATTCGCCGCTGGATTAAAGGACCAAG	248283
QY	538	CAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTTTGTATTATGCTGTTGGAGCGCTGTTA	597
Db	248284	CAGATGAAGGCGCTGAATAAAGCGTTTGGTTTGTATTATGCTGTTAGGCGGCTCCTG	248343
QY	598	GCATCGGCGAGGCATGCG	615
Db	248344	GCGTCGCCAAGACACCGG	248361

Search completed: March 28, 2004, 16:45:00  
Job time : 2577.83 secs

	Query Match	61.7%;	Score 379.2;	DB 1;	Length 258050;
	Best Local Similarity	78.0%;	Pred. No. 2e-109;		
	Matches 482;	Conservative	0;	Mismatches 133;	Indels 3; Gaps 2;
Qy	1	ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATGCATCATTTTAAACGCTGTGGCCA	60		
Db	247744	ATGACCTTTGAATGGTGGTTTGCGCTATCTACTGACCTCAACTCTGTGAGTCTTTCTCCG	247803		
Qy	61	GGCTCTGGTGCATCAACACATATGACCACCTCGCTCAACCAGGTTATCCGGCCG--GTG	118		

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 28, 2004, 12:14:43 ; Search time 306.314 Seconds  
(without alignments)  
8529.292 Million cell updates/sec

Title: US-09-847-392-1\_COPY\_557\_1171  
Perfect score: 615  
Sequence: 1 atgaccttagaattggtgtt.....tagcatcgaggcatggg 615

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	615	100.0	1200	3	Aaz94405 E. coli r
2	615	100.0	1231	3	Aaa48442 E. coli L
3	568.2	92.4	622	5	Aas93395 DNA encod
c	4	486.2	79.1	2216	5 AAs89415 DNA encod
	5	414	67.3	417	4 AAh84646 E. coli g
	6	414	67.3	417	7 ACa18626 Prokaryot
	7	379.2	61.7	621	7 ACa51728 Prokaryot
8	375	61.0	618	7	ACA36103 Prokaryot
9	357.4	58.1	618	7	ACA32157 Prokaryot
10	344.8	56.1	623	7	ACA48809 Prokaryot
c	11	274.8	44.7	621	7 ACA53970 Prokaryot
	12	259	42.1	259	7 ACA13399 Prokaryot
13	248.6	40.4	1884	5	AAs89412 DNA encod
c	14	207.4	33.7	263	4 AAh84479 E. coli g
	15	175	28.5	274	5 AAs89410 DNA encod
16	167.8	27.3	516	5	AAs93398 DNA encod
17	149.8	24.4	618	7	ACA52910 Prokaryot
18	141	22.9	1041	5	AAs93397 DNA encod
19	112.4	18.3	630	7	ACA42760 Prokaryot
20	104	16.9	630	7	ACA45284 Prokaryot
21	100	16.3	100	7	ACD71734 E. coli K
22	100	16.3	100	7	ACD71736 E. coli K
23	100	16.3	100	7	ACD71735 E. coli K

24	86.4	14.0	633	7	ACA43829 Prokaryot
25	80	13.0	627	7	ACA27048 Prokaryot
26	75.4	12.3	663	8	ADA29490 DNA encod
27	64.6	10.5	543	7	ACA20819 Prokaryot
28	58.8	9.6	533	7	ACA23772 Prokaryot
c	29	42.4	536	9	ADb68842 Minoroty
	30	39.8	536	6	ABK94661 G protein
	31	39.8	536	9	ADC51928 Mouse TGR
	32	39.8	536	6	ABK94660 G protein
33	39.8	6.5	1083	6	ABK94660 G protein
c	34	37.2	6.0	533	8 ADA31395 DNA encod
	35	36.8	6.0	1377	4 ABL28479 Drosophil
	36	36.8	6.0	5313	4 ABL28478 Drosophil
	37	36.6	6.0	2000	7 ADA71938 Rice gene
38	35.4	5.8	1028	4	AAF79862 Soybean c
39	34.6	5.6	357	6	ABK77410 Bacillus
40	34.4	5.6	283	6	ABN16733 Human ORF
41	33.8	5.5	987	6	ABK94565 G protein
42	33.8	5.5	987	7	ABT33314 GPR7 liga
43	33.8	5.5	987	8	ACC79865 Rat TGR26
44	33.8	5.5	987	9	ADC51916 Rat GPR8-
c	45	33.8	3292	9	ADB75566 Prostate

ALIGNMENTS

RESULT 1  
AAZ94405  
ID AAZ94405 standard; DNA; 1200 BP.  
XX  
AC AAZ94405;  
XX  
AC AAZ94405;  
DT 18-JUL-2000 (first entry)  
XX  
DE E. coli rhtB gene imparting homoserine resistance.  
XX  
KW Homoserine resistance; rhtB gene; L-homoserine; L-alanine; L-isoleucine;  
KW L-valine; L-threonine; ss.  
XX  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT CDS 557..1171  
FT /\*tag= a  
FT /note= "a DNA sequence corresponding to the coding region  
FT (minus the stop codon) is specifically claimed in Claim  
FT 3"  
XX EP994190-A2.  
XX  
PD 19-APR-2000.  
XX  
PP 20-SEP-1999; 99EP-00118581.  
XX  
PR 13-OCT-1998; 98RU-00118425.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Livshits VA, Zakataeva NP, Aleoshin VV, Belareova AV;  
PI Tokhmakova IL;  
XX  
DR WPI; 2000-273530/24.  
DR P-PSDB; AAY79298.  
XX  
PT Novel RhtB protein, useful for generation of L-homoserine resistance in  
PT Escherichia bacteria and large-scale production of e.g. L-homoserine and  
PT L-alanine.  
XX  
PS Claim 3; Page 10-11; 14pp; English.  
XX  
CC This is the DNA sequence the novel rhtB gene of Escherichia coli K-12.  
CC The rhtB gene maps at 86 min on the E. coli chromosome. The gene

CC participates in resistance to homoserine. Amplification of the gene  
CC results in an improvement of the amino acid productivity of E. coli. The  
CC invention provides: the RhtB protein (see AAY79298); DNA encoding the  
CC RhtB, especially nucleotides 557-1171 of the present sequence; a  
CC bacterium, especially of the genus Escherichia, in which L-homoserine  
CC resistance is enhanced by amplifying the copy number or increasing the  
CC expression rate of the rhtB DNA, the DNA being carried on a multicopy  
CC vector or on a transposon; and a method for producing an amino acid by  
CC cultivating the bacterium in a culture medium to produce and accumulate  
CC the amino acid in the medium, from which it is recovered. The method is  
CC used for the production of L-homoserine, L-alanine, L-isoleucine, L-  
CC valine or L-threonine (all claimed)

XX  
SQ Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 U; 0 Other;  
Query Match 100.0%; Score 615; DB 3; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 2e-193;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 60  
Db 557 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 616  
QY 61 GGCTCTGGTGCATCAACACATATGACACCTCGCTCAACACACGGTTATCCGGCCGGTGGC 120  
Db 617 GGCTCTGGTGCATCAACACATATGACACCTCGCTCAACACACGGTTATCCGGCCGGTGGC 676  
QY 121 GTCATTTGCTGGGTTTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 180  
Db 677 GTCATTTGCTGGGTTTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 736  
QY 181 GGGACGCTATTTCCCGCTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 240  
Db 737 GGGACGCTATTTCCCGCTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 796  
QY 241 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGCTGCAATTGACCTTAAATCG 300  
Db 797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGCTGCAATTGACCTTAAATCG 856  
QY 301 CTGGCCTCTACTCAATCGGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACC 360  
Db 857 CTGGCCTCTACTCAATCGGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACC 916  
QY 361 AATCCCAAAGTATTGTGTTTCTGGCGCGCTATTTCCGCAATTTCATGCGCGCAACAG 420  
Db 917 AATCCCAAAGTATTGTGTTTCTGGCGCGCTATTTCCGCAATTTCATGCGCGCAACAG 976  
QY 421 CCGCAACTGATGAGTATATCTGCTCGGCTCACCACCTATTTGCTGATATTTATTTG 480  
Db 977 CCGCAACTGATGAGTATATCTGCTCGGCTCACCACCTATTTGCTGATATTTATTTG 1036  
QY 481 ATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTCTATGATTAAAGGACCAAGCAG -540  
Db 1037 ATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTCTATGATTAAAGGACCAAGCAG 1096  
QY 541 ATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCA 600  
Db 1097 ATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCA 1156  
QY 601 TCGGCGAGGCATGCG 615  
Db 1157 TCGGCGAGGCATGCG 1171

RESULT 2  
AAA48442  
ID AAA48442 standard; DNA; 1231 BP.  
XX  
AC AAA48442;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE E. coli L-homoserine resistance gene, rhtB.

XX L-homoserine resistance; L-homoserine synthesis; rhtB; L-threonine;  
KW L-valine; L-leucine; ds.  
XX  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT CDS 557..1174  
FT /\*tag= a  
FT /product= "RhtB"  
XX  
PN EP1013765-A1.  
XX  
PD 28-JUN-2000.  
XX  
PF 20-DEC-1999; 99EP-00125406.  
XX  
PR 23-DEC-1998; 98RU-00123511.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
PI Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;  
XX  
XX WPI; 2000-414602/36.  
DR P-PSDB; AAY99597.  
XX  
PT Novel Escherichia bacterium having enhanced L-threonine resistance due to  
PT enhanced RhtC protein activity, used to produce L-threonine, L-  
PT homoserine, L-valine and L-leucine.  
XX  
PS Claim 4; Page 12-13; 24pp; English.  
XX  
CC The present sequence is the L-homoserine resistance gene, rhtB, from  
CC Escherichia coli. This sequence may be used to impart L-homoserine  
CC resistance on E. coli bacteria, which would be useful for producing a  
CC high yield of L-homoserine. L-homoserine resistance means that the  
CC bacteria will be able to grow on a minimal medium containing L-homoserine  
CC at a concentration at which the corresponding wild-type strain would not  
CC grow. Since the transformed bacteria can grow on the minimal medium, it  
CC can synthesize L-homoserine, which accumulates. The accumulated amino  
CC acids can then be removed from the culture medium. The bacterium of the  
CC present invention may also be used to synthesize L-threonine, L-valine  
CC and L-leucine at increased levels

XX  
SQ Sequence 1231 BP; 291 A; 319 C; 311 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 615; DB 3; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 2e-193;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 60  
Db 557 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 616  
QY 61 GGCTCTGGTGCATCAACACACTATGACACCTCGCTCAACACACGGTTATCCGGCCGGTGGC 120  
Db 617 GGCTCTGGTGCATCAACACACTATGACACCTCGCTCAACACACGGTTATCCGGCCGGTGGC 676  
QY 121 GTCATTTGCTGGGTTTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 180  
Db 677 GTCATTTGCTGGGTTTCAGACCGGACTGGCGATTTCATATTTGCTGGTGGCGGTT 736  
QY 181 GGGACGCTATTTCCCGCTCAGACCGGACTGGCGTTTGAAGTGTGAAGTGGGAGCGCGGCT 240  
Db 737 GGGACGCTATTTCCCGCTCAGACCGGACTGGCGTTTGAAGTGTGAAGTGGGAGCGCGGCT 796  
QY 241 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCG 300  
Db 797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCG 856  
QY 301 CTGGCCTCTACTCAATCGGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACC 360  
Db 857 CTGGCCTCTACTCAATCGGTCGACATTTTGTCCAGCGCGCAGTTTGTGTAATCTCACC 916

QY 361 AATCCAAAGTATTGTGTTCTGGCGGCGCTATTTCGGCAATTCATCGCCGCAACAG 420  
Db 917 AATCCAAAGTATTGTGTTCTGGCGGCGCTATTTCGGCAATTCATCGCCGCAACAG 976  
QY 421 CCGCAACTGATGCAGTATATCGTGTGCGGCGTCACCACTATTGTGGTCGATATTATTG 480  
Db 977 CCGCAACTGATGCAGTATATCGTGTGCGGCGTCACCACTATTGTGGTCGATATTATTG 1036  
QY 481 ATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGAATTAAGGACCAAGCAG 540  
Db 1037 ATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGAATTAAGGACCAAGCAG 1096  
QY 541 ATGAAGCGCTGAATAAGATTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTTAGCA 600  
Db 1097 ATGAAGCGCTGAATAAGATTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTTAGCA 1156  
QY 601 TCGGCGAGGCATGCG 615  
Db 1157 TCGGCGAGGCATGCG 1171

RESULT 3

AAS93395  
ID AAS93395 standard; cDNA; 622 BP.

AC AAS93395;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #29199.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG29208.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID NO 29199; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 622 BP; 123 A; 152 C; 172 G; 175 T; 0 U; 0 Other;

Query Match 92.4%; Score 568.2; DB 5; Length 622;  
Best Local Similarity 98.1%; Pred. No. 5.2e-178;  
Matches 607; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAA-CGCTCGGCC 59

Db 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAAGCCCGTCCCC 60

QY 60 AGGCTCTGGTCAATCAACACTATGACCACTCGCTCAACCCACGGTTAT--CCGCGCGGT 117

Db 61 AGGCTCTGGTCAATCAACACTATGACCACTCGCTCAACCCACGGTTATCGCGCGCGGT 120

QY 118 GCGCTCTATTGCTGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTTGGCGTGGG 177

Db 121 GCGCTCTATTGCTGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTTGGCGTGGG 180

QY 178 GTT-GGGAGCGCTATTTCCCGCTCAGTGATTGCGTTGAAGTGTGAAGTGGCAGCGCG 236

Db 181 GTTGGGAGCGCTATTTCCCGCTCAGTGATTGCGTTGAAGTGTGAAGTGGCAGCGCG 240

QY 237 GGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAA 296

Db 241 GGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAA 300

QY 297 ATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGAGTTTTTGTGAATCT 356

Db 301 ATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGAGTTTTTGTGAATCT 360

QY 357 CACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCATCGCCGCA 416

Db 361 CACCAATCCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCATCGCCGCA 420

QY 417 ACAGCCGCAACTGATGCGAGTATATCGTGCTCGGCTCACCACCTATTGTTGTCGATATTAT 476

Db 421 ACAGCCGCAACTGATGCGAGTATATCGTGCTCGGCTCACCACCTATTGTTGTCGATATTAT 480

QY 477 TGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAAGACCAAA 536

Db 481 TGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAAGACCAAA 540

QY 537 GCAGATGAAGCGCTGAATAAGATTTCGGCTCGTTGTTTATGTTGGTGGAGCGCTGTT 596

Db 541 GCAGATGAAGCGCTGAATAAGATTTCGGCTCGTTGTTTATGTTGGTGGAGCGCTGTT 600

QY 597 AGCATCGCGGAGGCATGCG 615

Db 601 AGCATCGCGGAGGCATGCG 619

RESULT 4

AAS89415/C

ID AAS89415 standard; cDNA; 2216 BP.

XX AAS89415;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #25219.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.



OS Homo sapiens.  
XX WO200175067-A2.  
PN 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US008631.  
PD 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
PF (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
XX P-PSDB; ABG25228.  
DR  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 25219; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2216 BP; 518 A; 562 C; 639 G; 497 T; 0 U; 0 Other;

Query Match 79.1%; Score 486.2; DB 5; Length 2216;  
Best Local Similarity 91.0%; Pred. No. 1.8e-150;  
Matches 563; Conservative 0; Mismatches 48; Indels 8; Gaps 4;

QY 5 CCTAGAAATGGTGGTTTGGCTACCTGCTGACATCGATCATTTTAAACGCTGCGCAGGCT 64  
Db CTTAGAAAGGGGGTTGCCACCCCGGACATTTGATCTTTTGTAGCCCGTTCCCGAGTTT 2025

QY 65 CTGGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CCGGCCGGTGGCGT 122  
Db CTGGCCCAATCACCACACTATGCCCCCTCGCTCAACACCGGTTATCGCGGCGCGGTGGCT 1965

QY 123 CTATTGCTGGGCTTCAGACCGGACTGGCGATTATATGTTGCTGGTGGCGTGGGTT-G 181  
Db CTATTGCTGGGCTTCAGACCGGACTGGCGATTATATGTTGCTGGTGGCGTGGGTTGG 1905

QY 182 GGACGCTATTTTCCCGCTCAGTGATTGGCTTTGAAGTGTGAAGTGGCGCAGGCGGGCTT 241  
Db GGACGCTATTTTCCCGCTCAGTGATTGGCTTTGAAGTGTGAAGTGGCGCAGGCGGGCTT 1845

QY 242 ACTTGATTGGCTGGGAATCAGCAGTGGCGCGCGCTGGTGGTGAATGACCTTAATCGC 301  
Db ACTTGATTGGCTGGGAATCAGCAGTGGCGCGCGCTGGTGGTGAATGACCTTAATCGC 1785

QY 302 TGGCCTCTACTCAATCGCTGACATTTGTTCCAGCG---CGAGTTTTTGTGAATCTC 357  
Db TGGCCTCTACTCAATCGCTGACATTTGTTTCTACAGCGCACATTTTGAATCTCTC 1725

QY 358 ACCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATCATCATGCCGCAA 417  
Db ACCATTCCAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATCATCATGCCGCAA 1665

QY 418 CAGCCGCAACTGATGATGATATATCGT-GCTCGGCGCTCACCACATTTGTGGTCGATATTAT 476  
Db CAGCCGCAACTGATGATGATATATCGTGGCTCGGCGCTCACCACATTTGTGGTCGATATTAT 1605

QY 477 TGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAA 536  
Db TGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAA 1545

QY 537 GCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTT 596  
Db GCAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTT 1485

QY 597 AGCATCGGCGGAGGCATGCG 615  
Db AGCATCGGCGGAGGCATGCG 1466

RESULT 5  
AAH84646  
ID AAH84646 standard; DNA; 417 BP.  
XX  
AC AAH84646;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE E. coli growth and proliferation related coding sequence SEQ ID NO:274.  
XX  
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;  
KW bacterial infection; microorganism; ds.  
XX  
OS Escherichia coli.  
XX  
PN WO200134810-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US030950.  
XX  
PR 09-NOV-1999; 99US-0164415P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Forsyth RA, Ohlsen K, Zyskind J;  
XX  
DR WPI; 2001-335933/35.  
XX P-PSDB; AAG98975.  
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
PT for screening for homologous genes and for designing expression vectors.  
XX  
PS Claim 9; Page 351; 522pp; English.  
XX  
CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli growth  
CC and proliferation related proteins given in AAG99078 and AAG98830 to  
CC AAG98999. (I) can be used as potential targets for the generation of new  
CC antimicrobial agents, and for identification of compounds which interact  
CC with the gene products of (I). In addition the expression of (I) and the  
CC purification of the proteins, the purified proteins can be used to  
CC generate reagents and screen small molecule libraries or other candidate  
CC compound libraries for compounds that can be further developed to yield  
CC novel antimicrobial compounds. In addition, nucleic acid probes  
CC complementary to (I) that are specific for particular species of  
CC microorganisms can be used to identify particular microorganism species  
CC in clinical specimens, therefore, providing a rapid and dependable method

by which to identify the causative agents of a bacterial infection. Also, antibodies generated against proteins translated from mRNA transcribed from proliferation-required sequences can also be used to screen for specific microorganisms that produce such proteins in a species-specific manner. AAH84371 and AAH84670 represent sequencing primers used in the isolation of *E. coli* growth and proliferation related sequence, which are used in an example from the present invention

Query Match	67.3%;	Score 414;	DB 4;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 7.6e-127;		

[illegible]

RESULT 6  
ACA18626  
ID ACA18626 standard; DNA; 417 BP.  
XX  
XX ACA18626;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
XX  
DE Prokaryotic essential gene #283.

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU14756.

Query Match	67.3%	Score 414;	DB 7;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 7.6e-127;		
Matches 414;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	202	GTGATTGGCTTTGAAGTGTGAAGTGGGCGAGGCGCGCTTACTTGATTGCTGGGAATC	261
Db	1	GTGATTGGCTTTGAAGTGTGAAGTGGGCGAGGCGCGCTTACTTGATTGCTGGGAATC	60
QY	262	CAGCAGTGGCGCGCGGTGGTGCAAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT	321
Db	61	CAGCAGTGGCGCGCGGTGGTGCAAATTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT	120
QY	322	CGACATTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAGTATTGTGTTT	381
Db	121	CGACATTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAAGTATTGTGTTT	180
QY	382	CTGGCGGCGCTATTTCCGCAATTTCATCATGCGCGCAACAGCGGCAACTGATGCAGTATATC	441
Db	181	CTGGCGGCGCTATTTCCGCAATTTCATCATGCGCGCAACAGCGGCAACTGATGCAGTATATC	240
QY	442	GTGCTCGGCGTCAACACTATTGTGGTGCATATTATTGTGATGATCGGTTACGCCACCCCTT	501
Db	241	GTGCTCGGCGTCAACACTATTGTGGTGCATATTATTGTGATGATCGGTTACGCCACCCCTT	300
QY	502	GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT	561
Db	301	GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT	360

QY 562 TTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCATCGCGGAGGCATGCG 615  
DB 361 TTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTAGCATCGCGGAGGCATGCG 414

RESULT 7  
ACA51728  
ID ACA51728 standard; DNA; 621 BP.  
AC ACA51728;  
XX 19-JUN-2003 (first entry)  
DE Prokaryotic essential gene #33385.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Salmomella typhi.

OS WO200277183-A2.  
PN 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU47858.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 39598; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 621 BP; 121 A; 170 C; 171 G; 159 T; 0 U; 0 Other;  
Query Match 61.7%; Score 379.2; DB 7; Length 621;  
Best Local Similarity 78.0%; Pred. No. 3.5e-115;  
Matches 482; Conservative 0; Mismatches 133; Indels 3; Gaps 2;

QY 1 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGACATCGATCATTTTACGCTGCGCA 60  
DB 1 ATGACCTTTGAATGGTGGTTTCCTATCTACTGACCTCAACTCTGTTGAGTCTTTCTCCG 60  
QY 61 GGCTCTGGTGCATCAACACTATGACCACTGCTCAACACGCTTATCCGGCGG--GTG 118  
DB 61 GGTTCAGGCGCCATCAATACCATGACGACGCTATCAACCATGGATATCGTGGCGCAGCG 120  
QY 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGCGTGGG 178  
DB 121 GCTTCTATCGCGGCTCCAGACCGGCTGCGGATACATATCGTACTGCTGGCGTCCGA 180  
QY 179 TTGGG-ACGCTATTTCCCGCTCAGTGAATGCGGTTTGAAGTGTGAAAGTGGCGAGCGCG 237  
DB 181 CTGGGTACGCTCTTTTCGCGCTCGCTCATCGCTTTTGAATTTCTGAAATGCGGCTGGCGCG 240  
QY 238 GCTTACTTGAATTTGGTGGGAATCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAA 297  
DB 241 GCTTATCTTATCTGGTGGTATCCAGCAATGGCGCGCGCTATCGATCTGCAT 300  
QY 298 TCGTGGCGCTCTACTCAATCGGCTCAGATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 357  
DB 301 ACTCTCGCCAGACGCAATCGCGGGTGGCTGTTTCAACCGGCGATATTGTCAATCTA 360  
QY 358 ACCAATCCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAATTCATCATGCCGCA 417  
DB 361 ACCAATCCCAAAAGTATTGTTTCTTGGCGCGCTGTTTCCGCAATTCATCATGCCGCA 420  
QY 418 CAGCGCAACTGATGTCAGTATATCGTCTCGGCTCAGCGCTCACCCTATTTGTTGCTGATATT 477  
DB 421 CAACCGCAACTGGCGCAGTACCTCATTTCTGGCGCTCACCAGATTGTTGGTGGATATGATT 480  
QY 478 GTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG 537  
DB 481 GTGATGACCGGTTACGCCACCGCTGGCGCAGCGCATTTGCCGCTGGATTAAAGGACCAAG 540  
QY 538 CAGATGAAGGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTA 597  
DB 541 CAGATGAAGGCGCTGAATAAGCGTTTGGTTCGTTGTTTATGCTGGTGGCGGCTCCTG 600  
QY 598 GCATCGCGGAGGCATGCG 615  
DB 601 GCGTCGGCAAGACACGCG 618

RESULT 8  
ACA36103  
ID ACA36103 standard; DNA; 618 BP.  
XX  
AC ACA36103;  
XX 19-JUN-2003 (first entry)  
DT Prokaryotic essential gene #17760.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
DE drug design; gene.  
KW Klebsiella pneumoniae.  
XX OS  
PN WO200277183-A2.  
XX

QY	178	GTTGGACGCTATTTCCCGCTCAGTGAATCCAGCAGTGGCGCGCTGGTGAATGACCTTAA	297
Db	181	CTCGGCACTCTCTTCTCCCGCTCGGTGCTGGCCCTTTGAGGTGCTGAAATGGCGCGGCC	240
QY	238	GCTTACTGATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGAATGACCTTAA	297
Db	241	GCCTACCTGATCTGGCTGGGTATCCAGCAATGGCGGCGCAGGGGCGATCGACCTGAAT	300
QY	298	TCGCTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGATCTC	357
Db	301	ACGCTGGCGAAGCGCAGACGCGCGGCAAACTGTTCCAGCGCGCGTTCGTAAATCTC	360
QY	358	ACCAATCCCAAAAGTATTGTTGTTTCTGGCGCGCTATTTCCGCAATTCATCATCGCGAA	417
Db	361	ACCAATCCCAAAAGCATTTGTTTCTGGCGCGCTGTTCCGAGTTTATCCTGCCGCAC	420
QY	418	CAGCCGCAACTGATGCAGTATATCGTCTCGGCGCTCACCACTATTGTGGTCGATATT	477
Db	421	CAGCCGCAAGTGATGCAATATCTCGTCTGGGCGTCAACCACTATCGTCGTCGATCAT	480
QY	478	GTGATGATCGGTTACGCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG	537
Db	481	GTGATGATTGGCTATCGACCTTGGCGCAGCGCATTTCCGTCATGGATCAAAAGGCCAAAG	540
QY	538	CAGATGAAGGCGCTGAATAGATTTTCGGCTCGTTGTTTATGCTGGGAGCGCTGTTA	597
Db	541	CAGATGAAGGCGTTGAACAAAGTCTTTGGTTCACTGTTTATGCTGGTGGGCGCTGCTC	600
QY	598	GCATCGGCGAGGCATGC	614
Db	601	GCCTCCGCCGCCCATGC	617
RESULT 9			
ACA32157			
ID	ACA32157 standard; DNA; 618 BP.		
XX	ACA32157;		
XX	19-JUN-2003 (first entry)		
XX	Prokaryotic essential gene #13814.		
DE	Antisense; ds; prokaryotic essential gene; cell proliferation;		
XX	drug design; gene.		
KW	Enterobacter cloacae.		
XX	WO200277183-A2.		
OS	03-OCT-2002.		
XX	21-MAR-2002; 2002WO-US009107.		
PN	21-MAR-2001; 2001US-00815242.		
XX	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		
XX	(ELIT-) ELITRA PHARM INC.		
PA	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;		
XX	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
PI	WPI; 2003-029926/02.		
PI	P-PSDB; ABU28287.		
XX	New antisense nucleic acids, useful for identifying proteins or screening		
PT	for homologous nucleic acids required for cellular proliferation to		
PT	isolate candidate molecules for rational drug discovery programs.		
XX			

PS Claim 14; SEQ ID NO 20027; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 618 BP; 106 A; 176 C; 183 G; 153 T; 0 U; 0 Other;  
Query Match 58.1%; Score 357.4; DB 7; Length 618;  
Best Local Similarity 75.9%; Pred. No. 6.1e-108;  
Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY 1 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATCGATCATTTTACGGTGTGCGCA 60  
Db 1 ATGACCTTCGAATGGTGGTTGCTACCTGCTGACATCGATCATTTTCTCTCGG 60  
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTA--TCGGCGCGGTG 118  
Db 61 GGCTCGGGGCTATTAAACACCATGACCACTCCATCAATCAGGCTACCGCGTGGCGG 120  
QY 119 GCGTCTATTGCTGGGTTTCAGACCGGACTGGCGATTATGCTGTTGCGGT-GGG 177  
Db 121 GCATCGATTGCGGTTTTCAGACCGGCTGGGTATTATATGCTGTTGCTGGGATCGGT 180  
QY 178 GTTGGGACGTATTTCCGGCTCAGTCATGCGTTGAGTGTGAAGTGGCAGGCGCG 237  
Db 181 CTGGGACCTGTTCTCCGCTCCGCTCCGCTGAGGTGCTGAAATGGCGCGGCA 240  
QY 238 GCTTACTTGGTGGGAATCCAGAGTGGCGCGCGCTGGTGCAATTGACCTTAA 297  
Db 241 GCGTATCTGATTTGGCTCGGCATCCAGCAGTGGCGCGCGGAGGCTCCATCAACCTGAAT 300  
QY 298 TCGTGGCTCTACTCAATCGCGTGCATGTTGTTCCAGCGCGAGTTTTGTGAATCTC 357  
Db 301 ACGTGGCTCTGACGCAAAACCGCGCGCATCTGTTAAGCGTGGCGTATTCGTCAACCTG 360  
QY 358 ACCAATCCAAAAGTATGTTGTTTTCGGCGCGCTATTTCCGCAATTCATCGCGCAA 417  
Db 361 ACCAACCAGAGAGCATGTTGTTCTCCGCGCGCTGTTCCCGCAGTTTATCGTCCGCAT 420  
QY 418 CAGCGCACTGATGCAATATATGCTGCTGGCGGTGACCACTATGTTGTTGCTGATATT 477  
Db 421 CAGCCTCAGTGTGATGATGCTGTTGCTGGCGCGGACCACTATATATGCTGATATCATC 480  
QY 478 GTGATGATCGTTACGCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAG 537

Db 481 GTGATGATTGGTTACGCGACGCTGGCGAGCGAATTCCGGCGTGGATTAAAGGCGCTAAG 540  
QY 538 CAGATGAAGCGCTGAATAAGATTTCGGCTCGTTGTTTATGCTGTTGGGAGCGCTGTTA 597  
Db 541 CAGATGAAGCGCTGAATAAGTCTTTGGCTCGCTGTTTATGCTGTTGGCGGCTGCTT 600  
QY 598 GCATCGCGGAGGCATGC 614  
Db 601 GCGTCAGCGGCTCACGC 617

RESULT 10  
ACA48809  
ID ACA48809 standard; DNA; 623 BP.  
XX ACA48809;  
XX 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #30466.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX Salmonella paratyphi.  
XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU44939.

New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 36679; 1766pp; English.  
PS The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of







AC AAS89412;  
XX 13-FEB-2002 (first entry)  
DT DNA encoding novel human diagnostic protein #25216.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS WO200175067-A2.  
XX PN 11-OCT-2001.  
XX PD 30-MAR-2001; 2001WO-US008631.  
XX PF 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG25225.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 1; SEQ ID NO 25216; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 1884 BP; 495 A; 464 C; 479 G; 446 T; 0 U; 0 Other;  
SQ Query Match 40.4%; Score 248.6; DB 5; Length 1884;  
Best Local Similarity 96.6%; Pred. No. 1.6e-71;  
Matches 254; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 320 GTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATCCCAAGATTGTGT 379  
DB 1160 GTCGGACTATGTATATTGCGGCAGTTTGTGAATCTCACCATCCCAAGATTGTGT 1219  
QY 380 TTCTGGCGGCGCTATTTCCGCAATTCATCATGCCGCAACAGCCGCACTGATGAGTATA 439  
DB 1220 TTCTGGCGGCGCTATTTCCGCAATTCATCATGCCGCAACAGCCGCACTGATGAGTATA 1279  
QY 440 TCGTGCTCGGCGTCACCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGCCACC 499  
DB 1280 TCGTGCTCGGCGTCACCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGCCACC 1339

QY 500 TTGCTCAACGGATTGCTCTATGGATTAAAGACCAAGCAGATGAAGCGCTGAATAAGA 559  
DB 1340 TTGCTCAACGGATTGCTCTATGGATTAAAGACCAAGCAGATGAAGCGCTGAATAAGA 1399  
QY 560 TTTTCGGCTCGTTGTTTATGCTG 582  
DB 1400 TTTTCGGCTCGTTGTTTATGCTG 1422  
RESULT 14  
AAH84479/c  
ID AAH84479 standard; DNA; 263 BP.  
XX AAH84479;  
AC AAH84479;  
XX 26-SEP-2001 (first entry)  
DT E. coli growth and proliferation related DNA sequence SEQ ID NO:107.  
XX Escherichia coli; growth; proliferation; microbial; antimicrobial;  
KW bacterial infection; microorganism; ds.  
XX Escherichia coli.  
OS WO200134810-A2.  
XX PN 17-MAY-2001.  
XX PD 09-NOV-2000; 2000WO-US030950.  
XX PF 09-NOV-1999; 99US-0164415P.  
XX PR (ELIT-) ELITRA PHARM INC.  
XX PI Forsyth RA, Ohlsen K, Zyskind J;  
XX WPI; 2001-335933/35.  
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
PT for screening for homologous genes and for designing expression vectors.  
PT Claim 1; Page 126; 522pp; English.  
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli growth  
CC and proliferation related proteins given in AAG99078 and AAG98830 to  
CC AAG98999. (I) can be used as potential targets for the generation of new  
CC antimicrobial agents, and for identification of compounds which interact  
CC with the gene products of (I). In addition the expression of (I) and the  
CC purification of the proteins, the purified proteins can be used to  
CC generate reagents and screen small molecule libraries or other candidate  
CC compound libraries for compounds that can be further developed to yield  
CC novel antimicrobial compounds. In addition, nucleic acid probes  
CC complementary to (I) that are specific for particular species of  
CC microorganisms can be used to identify particular microorganism species  
CC in clinical specimens, therefore, providing a rapid and dependable method  
CC by which to identify the causative agents of a bacterial infection. Also,  
CC antibodies generated against proteins translated from mRNA transcribed  
CC from proliferation-required sequences can also be used to screen for  
CC specific microorganisms that produce such proteins in a species-specific  
CC manner. AAH84371 and AAH84670 represent sequencing primers used in the  
CC isolation of E. coli growth and proliferation related sequence, which are  
CC used in an example from the present invention  
XX Sequence 263 BP; 72 A; 60 C; 68 G; 63 T; 0 U; 0 Other;  
SQ Query Match 33.7%; Score 207.4; DB 4; Length 263;  
Best Local Similarity 94.3%; Pred. No. 2.9e-58;  
Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 305 CCTCTACTCAATCGCGTCGACATTGTTCCAGCGCGCAGTTTGTGATCTCACCATC 364

Db 263 CCCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATC 204  
QY 365 CCAAAAAGTATTGTGTTCTGCGCGCGCTATTTCGCAATTCAATCGCGCAACAGCCGC 424  
Db 203 CCAAAAAGTATTGTGTTCTGCGCGCGCTATTTCGCAATTCAATCGCGCAACAGCCGC 144  
QY 425 AACTGATGTCAGTATATCGTGTC-GGCGTCACCACTATTGTGTCATATTATTGTGATG 483  
Db 143 AACTGATGTCAGTATATCGTGTCGGCGTCACCACTATTGTGTCATATTATTGTGATG 84  
QY 484 ATC-GGTTACGCCACCCTTGCTCAACGGATTG--CTCTATGGATTAAAGCAACAGCAG 540  
Db 83 ATCGGGTTACGCCACCCTTGCTCAACGGGATGGCTCTAAGGGATTAAAGGCCCAACCAA 24  
QY 541 ATGAAGGCGCTGAATAAGATTTT 563  
Db 23 ATGAAGGCGCTGAATAAGATTTT 1

RESULT 15

AAS89410/c

ID AAS89410 standard; cDNA; 274 BP.

XX AC AAS89410;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #25214.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG25223.

XX FT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX PS Claim 1; SEQ ID NO 25214; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 274 BP; 61 A; 71 C; 70 G; 72 T; 0 U; 0 Other;

Query Match 28.5%; Score 175; DB 5; Length 274;

Best Local Similarity 100.0%; Pred. No. 1.7e-47;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 409 ATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGCGTCACCACTATTGTGGTC 468

Db 274 ATGCCGCAACAGCCGCAACTGATGCAGTATATCGTCTCGCGTCACCACTATTGTGGTC 215

Qy 469 GATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATGGATTAAA 528

Db 214 GATATTATTGTGATGATCGGTTACGCCACCCTTGCTCAACGGATTGCTCTATGGATTAAA 155

Qy 529 GGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGTCTGTTGTTTATGCTGG 583

Db 154 GGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGTCTGTTGTTTATGCTGG 100

Search completed: March 28, 2004, 14:38:03

Job time : 310.314 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:10:49 ; Search time 61.6694 Seconds  
(without alignments)  
5534.263 Million cell updates/sec

Title: US-09-847-392-1\_COPY\_557\_1171  
Perfect score: 615  
Sequence: 1 atgaccttagaattggtggtt.....tagcatggcgaggatgcg 615

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	615	100.0	1200	4	US-09-396-357-1
2	414	67.3	417	4	US-09-711-164-274
3	375	61.0	645	4	US-09-489-039A-895
C 4	207.4	33.7	263	4	US-09-711-164-107
5	112.4	18.3	834	4	US-09-252-991A-6447
C 6	102.4	16.7	930	4	US-09-252-991A-6103
7	75.4	12.3	663	4	US-09-328-352-777
8	59	9.6	645	4	US-09-489-039A-2622
9	37.2	6.0	633	4	US-09-328-352-2682
10	36	5.9	312	4	US-09-543-681A-3779
11	35.4	5.8	1028	4	US-09-655-908-17
12	32.8	5.3	1442	4	US-09-634-238-82
13	32.2	5.2	681	4	US-09-252-991A-15039
C 14	32	5.2	1001	4	US-09-641-638-278
C 15	32	5.2	1597	4	US-09-634-238-195
C 16	31.8	5.2	738	4	US-09-252-991A-6015
17	31.8	5.2	1128	4	US-09-252-991A-6376
18	31.8	5.2	2136	4	US-09-252-991A-6294
C 19	31.8	5.2	3948	4	US-09-252-991A-6175
20	31.2	5.1	505	4	US-09-621-976-15639
21	31	5.0	540	4	US-09-489-039A-5266
22	31	5.0	726	4	US-09-489-039A-2390
23	31	5.0	1179	4	US-09-489-039A-2656
C 24	30.6	5.0	351	4	US-09-543-681A-2733
C 25	30.6	5.0	1137	4	US-09-252-991A-6183
26	30.6	5.0	1698	4	US-09-489-039A-5833
C 27	30.4	4.9	1014	4	US-09-325-932A-15

ALIGNMENTS

RESULT 1

US-09-396-357-1  
; Sequence 1, Application US/09396357  
; Patent No. 6303348  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD  
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/396.357  
; CURRENT FILING DATE: 1999-09-15  
; EARLIER APPLICATION NUMBER: RU98118425  
; EARLIER FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (557)..(1171)  
US-09-396-357-1

Query Match		100.0%;	Score 615;	DB 4;	Length 1200;
Best Local Similarity		100.0%;	Pred. No. 7.3e-200;		
Matches		615;	Conservative	0;	Mismatches
				0;	Indels
				0;	Gaps
QY	1	ATGACCTTAGAATGGTGGTTCGCTACCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA	60		
Db	557	ATGACCTTAGAATGGTGGTTCGCTACCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA	616		
QY	61	GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGTTATCCGCCGCGTGGC	120		
Db	617	GGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCGTTATCCGCCGCGTGGC	676		
QY	121	GTCTATTGCTGGGCTTCAGACCGGACTGGCGGATTCATATTGTGCTGGTGGCGTGGGTT	180		
Db	677	GTCTATTGCTGGGCTTCAGACCGGACTGGCGGATTCATATTGTGCTGGTGGCGTGGGTT	736		
QY	181	GGGACGCTATTTCCCGCTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	240		
Db	737	GGGACGCTATTTCCCGCTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	796		
QY	241	TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCAATTGACCTTAAATCG	300		



Db 797 TACTTGATTGGCTGGGAATCCAGAGTGGCGCGCTGGTGAATGACCTTAAATCG 856  
QY 301 CTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGATCTCACC 360  
Db 857 CTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGATCTCACC 916  
QY 361 AATCCCAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCAATTTCATCATGCCGCAACAG 420  
Db 917 AATCCCAAAGTATTGTGTTCTGGCGGCGCTATTTCCGCAATTTCATCATGCCGCAACAG 976  
QY 421 CCGCAACTGATGCAGTATATCGTGTCTCGCGTCAACCACTATTGTGTCGATATTATTGTG 480  
Db 977 CCGCAACTGATGCAGTATATCGTGTCTCGCGTCAACCACTATTGTGTCGATATTATTGTG 1036  
QY 481 ATGATCGGTTACGCCACCCCTGCTCAACGAGTTGCTCTATGGATTAAAGGACCAAGCAG 540  
Db 1037 ATGATCGGTTACGCCACCCCTGCTCAACGAGTTGCTCTATGGATTAAAGGACCAAGCAG 1096  
QY 541 ATGAAGCGGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTAGCA 600  
Db 1097 ATGAAGCGGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTGGTGGAGCGCTGTAGCA 1156  
QY 601 TCGGCGAGGCATGCG 615  
Db 1157 TCGGCGAGGCATGCG 1171

RESULT 2  
US-09-711-164-274  
; Sequence 274, Application US/09711164  
; Patent No. 6589738  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
; FILE REFERENCE: ELITRA.008A  
; CURRENT APPLICATION NUMBER: US/09/711,164  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-9  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 274  
; LENGTH: 417  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(417)  
US-09-711-164-274

Query Match 67.3%; Score 414; DB 4; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.9e-131;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 202 GTGATTGCGTTGAAGTGTGAAGTGGGCGGCGGCTTACTTGAATTGGCTGGGAATC 261  
Db 1 GTGATTGCGTTGAAGTGTGAAGTGGGCGGCGGCTTACTTGAATTGGCTGGGAATC 60  
QY 262 CAGCAGTGGCGCGCGCTGGTGCATTAAGCTTAAATCGCTGGCCTCTACTCAATCGCGT 321  
Db 61 CAGCAGTGGCGCGCGCTGGTGCATTAAGCTTAAATCGCTGGCCTCTACTCAATCGCGT 120  
QY 322 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTACCAATCCCAAAAGTATTGTGTTT 381  
Db 121 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTACCAATCCCAAAAGTATTGTGTTT 180  
QY 382 CTGCGGCGCTATTTCCGCAATTCATCATGCGGCAACAGCGCAACTGATGCAGTATATC 441  
Db 181 CTGCGGCGCTATTTCCGCAATTCATCATGCGGCAACAGCGCAACTGATGCAGTATATC 240

QY 442 GTGCTCGGCGTCACCACCTATTGTGTCGATATTATGTCATGATCGGTACGCCACCCTT 501  
Db 241 GTGCTCGGCGTCACCACCTATTGTGTCGATATTATGTCATGATCGGTACGCCACCCTT 300  
QY 502 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 561  
Db 301 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360  
QY 562 TTCGGCTCGTTGTTTATGTCGGTGGAGCGCTGTTAGCATCGGCGAGGCATGCG 615  
Db 361 TTCGGCTCGTTGTTTATGTCGGTGGAGCGCTGTTAGCATCGGCGAGGCATGCG 414  
RESULT 3  
US-09-489-039A-895  
; Sequence 895, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 895  
; LENGTH: 645  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-895

Query Match 61.0%; Score 375; DB 4; Length 645;  
Best Local Similarity 77.6%; Pred. No. 5.1e-118;  
Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;  
QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAAGCTGTGCGCA 60  
Db 25 ATGACCATTAGAGTGGTGGTTTGCCTACCTGCTGACATCAATTATTCTCAGCCTGTGCGCG 84  
QY 61 GGCTCTGGTGCATCAACACTATGACCACCTGCTCAACCACGGTTA--TCCGGCCGGTG 118  
Db 85 GGTCCGGAGCGATTATACCATGACCACCTCCATTAAACCACGGATACCGCGGCGCGCG 144  
QY 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATATTTGCTGTTGGCG-TGGG 177  
Db 145 GCGTCGATTGCCGGTTTGCAGACTGGGCTGGCCATTATATCTGCTGCTGCTGCTGCT 204  
QY 178 GTTGGGACGCTATTTTCCCGCTCAGTGAITGCGTTTGAAGTGTGGAAGTGGGCGAGCGCG 237  
Db 205 CTCGCACTCTCTTCTCCCGCTCGGTGCTGCGCTTTAGTGCTGAATGGGCGCGCGCC 264  
QY 238 GCTTACTTGAATTGGCTGGGAATCCAGACGTGGCGCGCGCTGGTGCAATTGACCTTAAA 297  
Db 265 GCCTACCTGATCTGGCTGGGTATCCAGCAATGGCGGCGGCGAGGCGCATCGACCTGAAT 324  
QY 298 TCCTGGCCTCTACTCAATCGGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 357  
Db 325 ACCTGGCGAAGCGCAGACGCGCGGCAAACTGTTTTCAGCGCGCGGTGTTCTGTAATCTC 384  
QY 358 ACCAATCCCAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCTCATGCGCA 417  
Db 385 ACCAATCCCAAAGCATTTGTTTCTGGCGCCCTGTTTTCGCGCAGTTTATCTCTCGCGCAC 444  
QY 418 CAGCGCAACTGATGCAGTATATCGTGTCTCGGCGTCACCACTATTGTGTCGATATTATT 477  
Db 445 CAGCGCAGTGTGCAATATATCGTTCTGGGCGTCAACCACCATCGTCTCGATATCAT 504  
QY 478 GTGATGATCGGTTACGCCACCCCTTGTCTCAACGGAATTGCTTATGGATTAAAGGCCAAAG 537  
Db 505 GTGATGATTGGCTATGCGACCCCTGGCGGCGCATTTTCCGCATGGATCAAAAGGGCCAAAG 564

Qy	538	CAGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTTGTTTATCTCGTGGAGCGCTGTTA	597
Dδ	565	CAGATGAAGCGCGTTGGAACAAAGTCTTTGGTTCACTGTTTATCTGGTCGGGGCGCTGCTC	624
Qy	598	GCATCGGCGAGGCATGC	614
Dδ	625	GCCTCGCGCCGCCCATGC	641

```

RESULT 4
US-09-711-164-107/C
; Sequence 107, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-711-164-107

```

	Query Match	33.7%;	Score 207.4;	DB 4;	Length 263;	
	Best Local Similarity	94.3%;	Pred. No. 5.6e-61;			
	Matches 248;	Conservative	0;	Mismatches 11;	Indels 4;	Gaps 3;
QY	305	CCtCTACTCAATCGCGTCGACATTGTTCCAGCGCGCAGTTTTTGTGGAATCTCAACAATC	364			
Db	263	CCCCTACTCAATCGCGTCGACATTGTTCCAGCGCGCAGTTTTTGTGGAATCTCAACAATC	204			
QY	365	CCAAAGTATTGTCTTTCTGGCGCGCTATTTCGGCAATTTCATCATGCGCGCAACAGCCGC	424			
Db	203	CCAAAGTATTGTGTTTCTGGCGCGCTATTTCGGCAATTTCATCATGCGCGCAACAGCCGC	144			
QY	425	AAC TGATGCAGTATATCGTGCTC - GGCGTCACCACTATTGTGGTCGATATTATTGTGATG	483			
Db	143	AAC TGATGCAGTATATCGTGCTGGGCGTCACCACTATTGTGGTCGATATTATTGTGATG	84			
QY	484	ATC - GGTACGCCACCCTTGCTCAACGGATTG - CTCTATGGATTAAAGGACCAAGCAG	540			
Db	83	ATCGGGTTACGCCACCCTTGCTCAACGGGATGGCTCTAAGGGATTAAAGGCCCAACCAA	24			
QY	541	ATGAAGCGCGCTGAATAAGATTTT	563			
Db	23	ATGAAGCGCGCTGAATAAGATTTT	1			

RESULT 5  
US-09-252-991A-6447  
; Sequence 6447, Application US/09252991A  
; Patent NO. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/034,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6447

```

; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6447

Query Match      18.3%; Score 112.4; DB 4; Length 834;
Best Local Similarity 53.5%; Pred. No. 3.1e-28;
Matches 324; Conservative 0; Mismatches 276; Indels 6; Gaps 4;

QY 16 TGGTTTGCCTACCTGCTGACATCGATCATATTTAACGGTGTGCGCAGGCTCTGTGTGCAATC 75
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 TGGTTTCGCCCTTCTTCCTCGCCTGCTGGGCCATCAGCCTGTGCGCGGCGCGCGAGCCATC 279

QY 76 AACACTATGACACACCTCGCTCAACCAACCGTTATCCGGCCCGTGGC--GTCTATTGCTGGG 133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 GCCTCGATGTCCTCGGGTTGCAGTACGGCTTTGCGCGAGGCTACTGGAAACGGCTGGGC 339

QY 134 CTTTCAGACCGGACTGGCGATTTCATATTGTGCTGGTTGGCG-TGGGGTTGGGACGCTATTT 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 CTGCAGATCGGCTGGCCCTGCAGATCGCCATTGTGCCCGCGCGTCCGTGGTGGTGGCTG 399

QY 193 TCCCGCTCAGTGATTGCGTTTGAAGTTGAAGTGGGCAGGCGCGCTTACTTGATTGG 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 GCGACCTCGGCACTGGCTTTCAGCCTGATCAAGTGGTTGCGCGTGCCTACCTGGTGTAC 459

QY 253 CTGGAATCCAGCAGTGGC--GCGCGCTGGTGCAATTGACCTTAAATCGCTGGCCTCTA 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 CTGCGGTTGCGCAGTGGCAGGCGCGCACAGGCCCTTGAGCACCGATGGCGAACGGCCT 519

QY 311 CTCATTCGCTCGACATT-TGTTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAAA 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 520 CTGGGCGCACCGTTGACCTTGGTGCTGGCTGGTTTCCTGGTCAACGCCAGCAATCCCAAG 579

QY 370 AGTATTGTGTTCTGGGGGCGCTATTTCGCGCAATTTCATCATATGCCGCAACAGCCGCACTG 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 580 GCGGTGATCTTCATGCTCGCGGTGCTGCGCGAGTTTCATCGAACCGACCGCCGCTGCTG 639

QY 430 ATGCAGTATATCGTGCTCGGCGTCACCACTATTGTGGTCGATATTATTGTGATGATCGGT 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 640 GCGCAATACCTGATCATGGCGGCGCACCATGATCGTCTGCGACCTGATCGTCAATGCGCCGC 699

QY 490 TACGCCACCCCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGGCG 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 700 TACACGGGCTGGCTCGCGCGGTGCTACGCGTACTCGGTTTCGCGCGGCCAGCAGAAGCTG 759

QY 550 CTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTAGCATCGGCGAGG 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 760 GTGAACCGTACCTTCGCCAGCCTGTTCTGCGGTGCCCGGGGCTGCTGGCGACGGTACGC 819

QY 610 CATGCG 615
    |||||
Db 820 CGAGCG 825
    |||||

```

RESULT 6  
US-09-252-991A-6103/c  
; Sequence 6103, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6103  
; LENGTH: 930  
; TYPE: DNA

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6103

Query Match      16.7%; Score 102.4; DB 4; Length 930;
Best Local Similarity 55.6%; Pred. No. 8.8e-25;
Matches 278; Conservative 0; Mismatches 216; Indels 6; Gaps 4;

QY 16 TGGTTTGCCTACTGCTGACATCGATCAATTTAAACGCTGTGCGCAGGCTCTGGTGCAATC 75
Db 504 TGGTTGCGCTTCTCTCGCTGCTGGGCGATCAGCCTGTGCGCGCGCGCGGAGCCATC 445

QY 76 AACACTATGACACCTCGCTCAACCAACCGTTATCCGGCCGGTGGC--GTCTATTGCTGGG 133
Db 444 GCTCGATGTCCTCGGGTTTCAGTACCGGCTTTGCGCAGGCTACTGGAACGCGGTGGC 385

QY 134 CTTACAGACCGGACTGGGATTCATATTTGCTGTTGGCG--TGGGTTGGGACGCTATTT 192
Db 384 CTCAGATCGGCTGCGCCCTGCAGATCGCCATTTGCGCGCGCGGCTGCGTGGTCTG 325

QY 193 TCCGCTCAGTATGCGTTTGAAGTGTGAAGTGGGAGGCGCGGCTTACTTGAATTTGG 252
Db 324 GCGACCTCGGCACTGGCTTTTCAGCCTGATCAAGTGGTTGCGGCTGCGCTACCTGGTGAC 265

QY 253 CTGGGAATCCAGCAGTGGC--CGCGCGCTGGTGCAATTTGCAATCGCTTAAATCGCTGGCCTCTA 310
Db 264 CTGGCGGTGCGCAGTGGCAGGCGCGCCACAGGCCCTTGAGCACCGATGGCGACGGCCT 205

QY 311 CTCAATCG-CGTGACATTTTGTCCAGCGCGCAGTTTTTGTGAATCTCACCAATCCCAA 369
Db 204 CTGGGGCGACCGTTGACCTGTGCTGCTGGTGGTTTCTTGGTCAACGCGCAATCCCAAG 145

QY 370 AGTATTGTTTTCGCGGCGTATTTCCGCAATTCATCATGCGCGCAACAGCCGCAACTG 429
Db 144 GCGTGATCTTCATGCTCGCGTGCTGCGCAGTTTCATCGACCGCGCAGCGCGTCTG 85

QY 430 ATGCAGTATATCGTCTGCGGCTGCGGCTCACCACCTATTGTTGTCGATATTTATGATCGGT 489
Db 84 GCGCAATACCTGATCATGTTGGCGGCACCATGATCGTCTGCGACCTGATCGTATGCGCGGC 25

QY 490 TAGCCACCCCTTGCTCAACG 509
Db 24 TACACCGGCTGCTGCGCG 5

RESULT 7
US-09-328-352-777
; Sequence 777, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 777
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-777

Query Match      12.3%; Score 75.4; DB 4; Length 663;
Best Local Similarity 49.1%; Pred. No. 1.2e-15;
Matches 286; Conservative 0; Mismatches 291; Indels 6; Gaps 3;

QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 60
Db 37 ATGTCTTTGCAAGTTTGGTTTGTCTTAATGTTGGCTGTGGTCACTAGTATTTCTCCA 96

QY 61 GGCTCTGGTGCATCAACACTATGACCACCTCGCTCAACACCGGTTATCCGCGCGGTGGC 120
Db 97 GGAGCAGGGGCAATGCTTCTATGTCGAGCGGCTTAAATATATGGCTTAGACATGGCTAC 156
```

```
QY 121 --GTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGTGCTGGTGGCGTGGG- 177
Db 157 TGGAAATGCAATTTGGTTTGCAATCGCCTTGTAAATTCAAATCATGATTGTAGCGCGGT 216

QY 178 GTTGGGACGCTATTTCCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGGAGCGCG 237
Db 217 GTAGTGTTCATTTTCGCAACGACACCATTTAGCCTTTTCAGCGGTTAAGTGGTTTGGGTA 276

QY 238 GCTTACTTGAATTTGGCTGGGAATCCAGAGTGGCGCGCCGCTGGTGA--ATTGACCTT 294
Db 277 GCTTATTTATTGATTTTAGCCTATTTTGCAGTGGACAGCCTGTAAAGATATAGAAAT 336

QY 295 AAATCGCTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAAT 354
Db 337 CAACAGCAAAAAGAGATAAATCTGTTTCTGCTCTGCTGTTCAATGGATTGTAGTCAAT 396

QY 355 CTCACCAATCCCAAAAGTATTGTGTTTCTGCGGCGGCTATTTCCGCAATTCATCATGCCG 414
Db 397 ATCAGTAACCTAAAGCCATCGTATTTTATTTGGCAGTTTGGCCTCAGTCTTAGATTTA 456

QY 415 CAACAGCGCAACTGATGTCAGTATATCGTCTGCGGTCACCACCTATTTGTTGGTGCATATT 474
Db 457 AGCAACCTCAATGGATACATATCTGATTATGGCAGCCACTATGGTTAGGATTGTTG 516

QY 475 ATTGTGATGATCGGTTACGCCACCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCA 534
Db 517 ATTGTAATGGCTGTTATACAGGCTAGCTTCAAAAGTTTTAAGATTGCTACGTTCTCCT 576

QY 535 AAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTA 577
Db 577 AGCAACAAAATATTATAACCGTGGTTTTCAGTCAATGTTTA 619
```

```
RESULT 8
US-09-489-039A-2622
; Sequence 2622, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2622
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2622
```

```
Query Match      9.6%; Score 59; DB 4; Length 645;
Best Local Similarity 46.1%; Pred. No. 4.8e-10;
Matches 197; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 174 TGGGTTGGGACGCTATTTTCCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGGAGG 233
Db 204 TGGTTAGGTGCGCTACTGCGGTTTCTGAGTTGGCTTATACCTTATTGAAATGGTGGCG 263

QY 234 CGCGCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATGACCT 293
Db 264 AGCAGGATACCTTGTGCTGGTGGGATCCAGTTGCTCTCGACCAACGCCAGCAATTTAA 323

QY 294 TAAATCGCTGGCCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGAGTTTTTGTGA 353
Db 324 CACCCACCCGCGCAATCAGATTCACAGTAACCTGGTTCTGCTGGGATGCTGGGAA 383

QY 354 TCTCACCATCCCAAAAGTATGTGTTTCTGGCGCGCTATTTCCGCAATTCATATGCC 413
Db 384 TGTGCTTAACCCCAAAATGGCGCTCTTCTACGTTTTCGTTTACCTCAGTTATCCCGC 443
```

RESULT 10  
US-09-543-681A-3779  
; Sequence 3779, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706

RESULT 12  
US-09-634-238-82  
; Sequence 82, Application US/09634238  
; Patent No. 654772  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka J.

APPLICANT: Bloksberg, Leonard, N.  
APPLICANT: Lubbers, Mark W.  
APPLICANT: Dekker, James  
APPLICANT: Christensson, Anna C.  
APPLICANT: Holland, Ross  
APPLICANT: O'Toole, Paul W.  
APPLICANT: Reid, Julian R.  
APPLICANT: Coolbear, Timothy  
TITLE OF INVENTION: Polynucleotides, materials incorporating  
TITLE OF INVENTION: them and methods for using them.  
FILE REFERENCE: 11000.1043UI  
CURRENT APPLICATION NUMBER: US/09/634,238  
CURRENT FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 82  
LENGTH: 1442  
TYPE: DNA  
ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-82

Query Match 5.3%; Score 32.8; DB 4; Length 1442;  
Best Local Similarity 45.4%; Pred. No. 0.68;  
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 309 TACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCAATCCCAA 368  
Db 185 TAGACAGACTCGTCTGCTACTTTATAGGAGGAAGTTAAACATGCGTTAAACAATTACC 244  
QY 369 AAGTATTGTGTTTCTGGCGCGCTATTTCGCAATCATCATCGCGCAACAGCGCAACT 428  
Db 245 AAGGACTTTGTAATCGGTGGCGCACTGCTGTACCAAGTTGAAGGGGCAACCAAGAA 304  
QY 429 GATGCAGTATATCGTGTGGCGGTCCACACTATTGTGTCGATATTATTGTGATGATCGG 488  
Db 305 GACGGAAAGGTGCGAGTCTTTGGGATGATTTTCTGGAATAACAGGGCGGTTAGTCCT 364  
QY 489 TTACGCCACCCCTTGCTCAACGAGTTGCTCTATGGATTAAAGGACCAAGCAGATGAAGGC 548  
Db 365 GACCCCGCGCTGATTTTATCATCGTATGATGAGGATTTGGCGTTAGCAGACATAT 424  
QY 549 GCTGAATAAGATTTTCGGCT 568  
Db 425 GGTCAATCAAGTAATACGGCT 444

RESULT 13  
US-09-252-991A-15039  
Sequence 15039, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15039  
LENGTH: 681  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15039

Query Match 5.2%; Score 32.2; DB 4; Length 681;  
Best Local Similarity 57.4%; Pred. No. 0.71;  
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 175 GGGGTTGGGACGCTATTTCCCGCTCAGTGATTCGTTGAAGTGTGAAGTGGGCGAGGC 234

Db 241 GGCCTGGCGGTGCTGATCACCACCACCGCGTGGCTGTTCACGGCTGAAGTACACCGGC 300  
QY 235 GCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGGC 275  
Db 301 GCGGCTACTGATCTGGATCGGCATCCAGGCCCTGGCTC 341

RESULT 14  
US-09-641-638-278/c  
Sequence 278, Application US/09641638  
Patent No. 6432648  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM  
FILE REFERENCE: GENSET.051CP1  
CURRENT APPLICATION NUMBER: US/09/641,638  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: US 09/502,330  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: US 60/133,200  
PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: US 60/119,917  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm  
SEQ ID NO 278  
LENGTH: 1001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 12-835-54 : polymorphic base A or G  
NAME/KEY: misc\_binding  
LOCATION: 481..500  
OTHER INFORMATION: 12-835-54.mis1, potential  
NAME/KEY: misc\_binding  
LOCATION: 502..521  
OTHER INFORMATION: 12-835-54.mis2, potential complement  
NAME/KEY: primer\_bind  
LOCATION: 449..468  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer\_bind  
LOCATION: 879..898  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc\_binding  
LOCATION: 489..513  
OTHER INFORMATION: 12-835-54 potential probe  
NAME/KEY: misc\_feature  
LOCATION: 785  
OTHER INFORMATION: n=a, g, c or t  
US-09-641-638-278

Query Match 5.2%; Score 32; DB 4; Length 1001;  
Best Local Similarity 50.7%; Pred. No. 1;  
Matches 74; Conservative 1; Mismatches 71; Indels 0; Gaps 0;

QY 169 TGGCGTGGGTTGGACGCTATTTCCCGCTCAGTGATTCGTTGAAGTGTGAAGTGG 228  
Db 535 TGGGAGGGGGGGGATCTACTTTTACAACTCAGCATGTAGCAGATGAGTGAATTT 476  
QY 229 GCAGGCGCGGCTTACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATT 288  
Db 475 CCAGATGGTTCTCTTGATGACTGTTGGCATAATGSCAGTTTGGCCCTTGGGAGCATT 416  
QY 289 GACCTAAATCGCTGGCCTCTACTCA 314

Db 415 GGGCTGCAACCCCTGAAGCCTAGACA 390

RESULT 15

US-09-634-238-195/c  
; Sequence 195, Application US/09634238  
; Patent No. 6544772  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka J.  
; APPLICANT: Bloksberg, Leonard, N.  
; APPLICANT: Lubbers, Mark W.  
; APPLICANT: Dekker, James  
; APPLICANT: Christensson, Anna C.  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul W.  
; APPLICANT: Reid, Julian R.  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Polynucleotides, materials incorporating  
; TITLE OF INVENTION: them and methods for using them.  
; FILE REFERENCE: 11000.1043U1  
; CURRENT APPLICATION NUMBER: US/09/634,238  
; CURRENT FILING DATE: 2000-08-08  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 195  
; LENGTH: 1597  
; TYPE: DNA  
; ORGANISM: Lactobacillus rhamnosus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1597)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-634-238-195

Query Match 5.2%; Score 32; DB 4; Length 1597;  
Best Local Similarity 52.2%; Pred. No. 1.4;  
Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 361 AATCCCAAAAGTATTGTGTTCTGGCGCGCTATTTCGCAATTTCATCATGCCGCAACAG 420  
Db 440 AAACCGCAACGTTTACGCATGCGCTTTCGCGCAGCGCCCATGCGCTTGGG 381  
QY 421 CCGCAACTGATGCAGTATATCGTGTGGCGGTGACCACTATTGTGTCGATATTATTG 480  
Db 380 CAGCAAGGGAACCCGTGATTTTGTGGCAGTCGATCAACCGTGGTGGCCTAATCGCG 321  
QY 481 ATGATCGGTTACGCCA 496  
Db 320 TTGATGGATGAAGCCA 305

Search completed: March 28, 2004, 18:24:34  
Job time : 68.6694 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 16:45:05 ; Search time 261.248 Seconds  
(without alignments)  
8764.284 Million cell updates/sec

Title: US-09-847-392-1\_COPY\_557\_1171  
Perfect score: 615  
Sequence: 1 atgaccttagaatggtggtt.....tagcatcgaggagcatgcg 615

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	615	100.0	1200	9	US-09-927-395-1 Sequence 1, Appli
2	615	100.0	1200	9	US-09-847-392-1 Sequence 1, Appli
3	414	67.3	417	12	US-10-282-122A-6496 Sequence 6496, Ap
4	414	67.3	417	14	US-10-287-274-274 Sequence 274, App
5	379.2	61.7	621	12	US-10-282-122A-39598 Sequence 39598, A
6	375	61.0	618	12	US-10-282-122A-23973 Sequence 23973, A
7	357.4	58.1	618	12	US-10-282-122A-20027 Sequence 20027, A
8	344.8	56.1	623	12	US-10-282-122A-36679 Sequence 36679, A
9	274.8	44.7	621	12	US-10-282-122A-41840 Sequence 41840, A
10	259	42.1	259	12	US-10-282-122A-1269 Sequence 1269, Ap
11	207.4	33.7	263	14	US-10-287-274-107 Sequence 107, App
12	149.8	24.4	618	12	US-10-282-122A-40780 Sequence 40780, A
13	112.4	18.3	630	12	US-10-282-122A-30630 Sequence 30630, A
14	104	16.9	630	12	US-10-282-122A-33154 Sequence 33154, A
15	86.4	14.0	633	12	US-10-282-122A-31699 Sequence 31699, A

16	80	13.0	627	12	US-10-282-122A-14918	Sequence 14918, A
17	64.6	10.5	543	12	US-10-282-122A-8689	Sequence 8689, Ap
18	58.8	9.6	633	12	US-10-282-122A-11642	Sequence 11642, A
19	42.4	6.9	536	16	US-10-338-110-119	Sequence 119, App
20	39.8	6.5	987	12	US-10-433-561-139	Sequence 139, App
21	39.8	6.5	1083	12	US-10-433-561-137	Sequence 137, App
22	39.4	6.4	1049	12	US-10-142-426-358	Sequence 358, App
23	39.4	6.4	1049	14	US-10-123-155-358	Sequence 358, App
24	39.4	6.4	1049	14	US-10-146-731-358	Sequence 358, App
25	39.4	6.4	1049	14	US-10-140-472-358	Sequence 358, App
26	39.4	6.4	1049	14	US-10-141-761-358	Sequence 358, App
27	39.4	6.4	1049	14	US-10-142-885-358	Sequence 358, App
28	39.4	6.4	1049	14	US-10-158-790-358	Sequence 358, App
29	39.4	6.4	1049	15	US-10-137-871-358	Sequence 358, App
30	39.4	6.4	1049	15	US-10-140-923-358	Sequence 358, App
31	39.4	6.4	1049	15	US-10-141-756-358	Sequence 358, App
32	39.4	6.4	1049	15	US-10-141-759-358	Sequence 358, App
33	39.4	6.4	1049	15	US-10-140-805-358	Sequence 358, App
34	39.4	6.4	1049	15	US-10-140-864-358	Sequence 358, App
35	38.4	6.2	574	14	US-10-184-644-234	Sequence 234, App
36	38.4	6.2	574	14	US-10-184-634-234	Sequence 234, App
37	36.2	5.9	686	14	US-10-184-644-524	Sequence 524, App
38	36.2	5.9	686	14	US-10-184-634-524	Sequence 524, App
39	35.6	5.8	1049	12	US-10-142-426-358	Sequence 358, App
40	35.6	5.8	1049	14	US-10-123-155-358	Sequence 358, App
41	35.6	5.8	1049	14	US-10-146-731-358	Sequence 358, App
42	35.6	5.8	1049	14	US-10-140-472-358	Sequence 358, App
43	35.6	5.8	1049	14	US-10-141-761-358	Sequence 358, App
44	35.6	5.8	1049	14	US-10-142-885-358	Sequence 358, App
45	35.6	5.8	1049	14	US-10-158-790-358	Sequence 358, App

ALIGNMENTS

RESULT 1

US-09-927-395-1  
; Sequence 1, Application US/09927395  
; Patent No. US20020058314A1  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD  
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/927,395  
; CURRENT FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 09/396,357  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: RU98118425  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1

LENGTH: 1200  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (557)..(1171)  
US-09-927-395-1

Query Match 100.0%; Score 615; DB 9; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 4.3e-206;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 60  
|||||  
Db 557 ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCATTTTAACGCTGCGCCA 615



; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6496
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-6496

Query Match 67.3%; Score 414; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.5e-135;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 GTGATTGCGTTTGAAGTGTGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 261
Db 1 GTGATTGCGTTTGAAGTGTGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 60

Qy 262 CAGCAGTGGCGGCGGCTGTGGAATTGACCTTAAATCGTGGCCTCTACTCAATCGCGT 321
Db 61 CAGCAGTGGCGGCGGCTGTGGAATTGACCTTAAATCGTGGCCTCTACTCAATCGCGT 120

Qy 322 CGACATTTGTTCCAGCGCGAGTTTGTGAATCTCACCATAATCCCAAAAGTATTGTGTTT 381
Db 121 CGACATTTGTTCCAGCGCGAGTTTGTGAATCTCACCATAATCCCAAAAGTATTGTGTTT 180

Qy 382 CTGGCGGCGCTATTCCGCAATTCATCATGCGCAACAGCGCAACTGATGCAATATATC 441
Db 181 CTGGCGGCGCTATTCCGCAATTCATCATGCGCAACAGCGCAACTGATGCAATATATC 240

Qy 442 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTGTGATGATGCGGTACGCCACCCCTT 501
Db 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTGTGATGATGCGGTACGCCACCCCTT 300

Qy 502 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 561
Db 301 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360

Qy 562 TTCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGATGCG 615
Db 361 TTCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGATGCG 414

RESULT 4
US-10-287-274-274
; Sequence 274, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09

; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(417)
US-10-287-274-274

Query Match 67.3%; Score 414; DB 14; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.5e-135;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 202 GTGATTGCGTTTGAAGTGTGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 261
Db 1 GTGATTGCGTTTGAAGTGTGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 60

Qy 262 CAGCAGTGGCGGCGGCTGTGCAATTGACCTTAAATCGTGGCCTCTACTCAATCGCGT 321
Db 61 CAGCAGTGGCGGCGGCTGTGCAATTGACCTTAAATCGTGGCCTCTACTCAATCGCGT 120

Qy 322 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATAATCCCAAAAGTATTGTGTTT 381
Db 121 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATAATCCCAAAAGTATTGTGTTT 180

Qy 382 CTGGCGGCGCTATTTCGCAATTCATCATGCGCAACAGCGCAACTGATGCAATATATC 441
Db 181 CTGGCGGCGCTATTTCGCAATTCATCATGCGCAACAGCGCAACTGATGCAATATATC 240

Qy 442 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTGTGATGATGCGGTACGCCACCCCTT 501
Db 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTGTGATGATGCGGTACGCCACCCCTT 300

Qy 502 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 561
Db 301 GCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360

Qy 562 TTCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGATGCG 615
Db 361 TTCGGCTCGTTGTTTATGCTGTGGGAGCGCTGTAGCATCGCGGAGGATGCG 414

RESULT 5
US-10-282-122A-39598
; Sequence 39598, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26





; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 36679  
; LENGTH: 623  
; TYPE: DNA  
; ORGANISM: Salmonella paratyphi A  
US-10-282-122A-36679

Query Match 56.1%; Score 344.8; DB 12; Length 623;

Best Local Similarity 76.6%; Pred. No. 9e-111;

Matches 475; Conservative 0; Mismatches 137; Indels 8; Gaps 4;

QY	1	ATGACCTTAGAATGGTGGTTGCTTACCTGCTGACATGATCAATTTAAACGCTGCGCA	60
Db	1	ATGACCTTTGAATGGTGGTTGCTTACCTGCTGACCTCAACTCTGCTGAGTCTTCTCG	60
QY	61	GGCTCTGTCGAATCAACACTATGACCACTCGCTCAACCAAGTTATCCGGCGG	-GTG 118
Db	61	GGTTCAGGGCGCATCAATACCATGACGAGCTCTATCAACCATGATATCGTGGCGCAG	120
QY	119	GGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGTTGGTGGG	178
Db	121	GCTTCTATCGCGGACTCCAGACCGGACTGGGATACATATCGTACTGTTGGGCTCG	180
QY	179	TTGGG-ACGCTATTTTCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCGCG	237
Db	181	CTGGGTACGCTCTTTTCGCGCTCGCTCCTCGCTTTTGAATTCGAAATGGGCTGG	240
QY	238	GCTTACTTGAATTTGGGAAATCCAGCAGTGGCGCGCTGTTGCAATTTGCAATCTA	297
Db	241	GCTTATCTTATCTGGTGGTATCCAGCAATGGCG--CGCAGGCGCTATCGATCTGC	297
QY	298	TCGCTGGCCTCTACTCAATCGCGTCCGACATTTGTTCCAGCGCGCAGTTTGTGA	357
Db	298	ACTCTCGCCAGACGCAATCGCGGGTGGCTGTTCAAAACGGCGATATTTGTCAATCT	357
QY	358	ACCAATCCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAATTTCATCATGCCG	417
Db	358	ACCAATCCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAATTTCATCATGCCG	417
QY	418	CAGCCGCACTGATGATGATATCGTCTGGCGTCCACCACTATTGTGTCGATATT	477
Db	418	CAACCGCACTGGCGCAGTACCTTCTCGCGTCCACCACTATTGTGTCGATATT	477
QY	478	GTGATGATCGGTTACGCCACCTTGTCTCAACGG--ATTGCTCTATGGAATTAAG	535
Db	478	GTGATGACCGGTTACGCCACACTGGCGGTAGCGACTTCCGCTGGAATTAAGGAC	537
QY	536	AGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGAGCG	595
Db	538	AGCAGATGAAGCGCTGAATAAGCGTTGTTGTTGTTTATGCTGGTGGAGCGCTCC	597
QY	596	TAGCATCGGCGAGGCATGCG	615

Db 598 TGGCTCGGCAAGACACGCG 617  
RESULT 9  
US-10-282-122A-41840  
; Sequence 41840, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41840  
; LENGTH: 621  
; TYPE: DNA  
; ORGANISM: Yersinia pestis  
US-10-282-122A-41840

Query Match 44.7%; Score 274.8; DB 12; Length 621;

Best Local Similarity 67.5%; Pred. No. 4.9e-86;

Matches 416; Conservative 0; Mismatches 197; Indels 3; Gaps 2;

QY	1	ATGACCTTAGAATGGTGGTTGCTTACCTGCTGACATGATCAATTTAAACGCTGCGCA	60
Db	1	ATGACCTTAGAATGGTGGTTAAACCTATCTGCTGACAACTTATGATCTCTCTCTCC	60
QY	61	GGCTCTGTCGAATCAACACTATGACCACTCGCTCAACCAAGTTATC--CGGCCG	118
Db	61	GGCTCGGGGCGCATTAACCAATGAGCACTGCCATCAGCCATGTTACCCGTTGTT	120
QY	119	GGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGTTGGCGT	177
Db	121	GCTTCCATTGGCGGGTTTAACTGGGTTGCGGTTGCGGTTGATATCGTCTGTTGG	180
QY	178	GTTGGGACGCTATTTTCCGCTCAGTGAATTCGTTTGAAGTGTGAAGTGGGCGCG	237
Db	181	CTTGGTGGCTGGTCTCTCAATCTTACTGGCATTTGAAATATTAAAGTGGTGGTGG	240



QY 238 GCTTACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATGACCTTAAA 297  
Db 241 GCTTACTTAATTTGGCTTGGTATTTCAACAATGGCTGACCGCGTGGCTGACCTGCAT 300  
QY 298 TCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTC 357  
Db 301 GCACTGGCAAAACAGTATGCCAGCGGTAAGCTGTTTAAACGTGCGGTGTTTGTAACTG 360  
QY 358 ACCAATCCAAAAGTATTTGTTCTGGCGGCGCTATTTCCGCAATTCATCATGCGCAA 417  
Db 361 ACCAATCCAAAAGCAATTTTCTGCGCGCGTATTTCCACAAATTTGTACTGCCACAG 420  
QY 418 CAGCCGCAACTGATGACATATATCGTGTGCGCGTCACCACTATTGTGGTCGATATTAT 477  
Db 421 CAACCGCAGTGGCACAGTATTGATTTGGGCAGCACCACTGTGATTGTGATATTATC 480  
QY 478 GTGATGATCGGTTACGCCACCTTTGCTCAACGGATTGCTCTATGGATTAAAGACCAAAG 537  
Db 481 GTGATGATGGTTATGCCACTTTTGGCTACACAGTATGCCCCGTTGGATTAAATCGCCACAG 540  
QY 538 CAGATGAAGCGCGCTGAATAAGATTTTTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGTTA 597  
Db 541 CAAATGAACACTCTTGAACCGAATATTGGCGGCGCTATTTATGTTGATTGTGTCATTATTA 600  
QY 598 GCATCGCGCAGGCATG 613  
Db 601 GCCACAGCGCGCAAAG 616

RESULT 10

US-10-282-122A-1269/c  
; Sequence 1269, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1269  
; LENGTH: 259  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-282-122A-1269

Query Match 42.1%; Score 259; DB 12; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.1e-80;  
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 CTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATCCCA 367  
Db 259 CTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATCCCA 200  
QY 368 AAAGTATTGTTTCTGGCGCGCTATTTCGCCAATTCATCATGCCGCAACAGCCGCAAC 427  
Db 199 AAAGTATTGTTTCTGGCGCGCTATTTCGCCAATTCATCATGCCGCAACAGCCGCAAC 140  
QY 428 TGATGCAGTATATCGTGTCTGGCGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCG 487  
Db 139 TGATGCAGTATATCGTGTCTGGCGCGTCAACCACTATTGTGGTCGATATTATTGTGATGATCG 80  
QY 488 GTTACGCCACCCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGG 547  
Db 79 GTTACGCCACCCCTTGTCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGG 20  
QY 548 CGCTGAATAAGATTTTCGG 566  
Db 19 CGCTGAATAAGATTTTCGG 1

RESULT 11

US-10-287-274-107/c  
; Sequence 107, Application US/10287274  
; Publication No. US20030181408A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET  
; FILE REFERENCE: ELITRA.008DV1  
; CURRENT APPLICATION NUMBER: US/10/287,274  
; PRIOR FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: US 60/164415  
; PRIOR FILING DATE: 1999-11-09  
; PRIOR APPLICATION NUMBER: US 09/711164  
; PRIOR FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107  
; LENGTH: 263  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-287-274-107

Query Match 33.7%; Score 207.4; DB 14; Length 263;  
Best Local Similarity 94.3%; Pred. No. 2e-62;  
Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 305 CCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATC 364  
Db 263 CCCCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCACCAATC 204  
QY 365 CCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGC 424  
Db 203 CCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTCATCATGCCGCAACAGCCGC 144  
QY 425 AACTGATGCAGTATATCGTGCTC-GGCGTCACCACTATTGTGGTCGATATTATTGTGATG 483  
Db 143 AACTGATGCAGTATATCGTGCTGGGCGTCACCACTATTGTGGTCGATATTATTGTGATG 84  
QY 484 ATC-GGTTACGCCACCCCTTGTCTCAACGGATTG-CTCTATGGATTAAAGGACCAAGCAG 540

```
Db      83 ATCGGGTACGCCACCCCTGCTCAACGGGATGGCTCTAAGGGATTAAAGGCCCAACCAA 24
QY      541 ATGAAGGCGCTGAATAAGATTTT 563
Db      23 ATGAAGGCGCTGAATAAGATTTT 1

RESULT 12
US-10-282-122A-40780
; Sequence 40780, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40780
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Vibrio cholerae
US-10-282-122A-40780

Query Match      24.4%; Score 149.8; DB 12; Length 618;
Best Local Similarity 55.5%; Pred. No. 7.3e-42;
Matches 330; Conservative 0; Mismatches 262; Indels 3; Gaps 2;

QY      16 TGGTTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCAGGCTCTGGTGCAATC 75
Db      16 TGGCTTGCTATCTACTGACCGCGGTGGTGTAGTCTGGCGCCGGTTTCAGGCACGGTT 75

QY      76 AACACTATGACCACCTCGCTCAACACACGGT--TATCCGGCCGGTGGCGTCTATTGCTGGG 133
Db      76 AACTCGATCAGCAATGGTTTAAGTTATGGTTACACGCCACTCATTAGGGCGGATTATCGGC 135

QY      134 CTTACAGCCGGAAGTGGCGATTATGTTGCTGGTGGC-GTGGGGTTGGACCGTATT 192
Db      136 CTACAAATCGGTCTTGGCTGTGTCATATAGTTGTGTCGGCATTCGAATTGGCGCTTTGGTG 195
```

```
QY      193 TCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGAGGCGCGGCTTACTTGAITGG 252
Db      196 GCACAAATCTGCGTTGGCGTTCACTCTGTATCAATGGATTGGCGGCTTATTTAGTATGG 255

QY      253 CTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTACCTTTAAATCGCTGGCCTCTACT 312
Db      256 CTGGGTATTAGAAATGGCGAGATAGAGCCCCCTTAACGGCTACCCACAACTTCTCATGAA 315

QY      313 CAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTGTGAATCTCAACCAATCCCAAAAGT 372
Db      316 CTCTCACAAGCCGCCCTGCTGCTGTAAGCAGTACTGATTAACCTGACCAATCCCAATCA 375

QY      373 ATTGTGTTTCTGGCGCGCTATTTCCGCAATTCATCATGCCGCAACAGCCGCAACTGATG 432
Db      376 ATCGTTTTCTTAGTGGCATTTATCCGCAGTTTATTGATCCCACTAGAGATCATTTGCCCG 435

QY      433 CAGTATATCGTGTCTGCGGCTCACCACCTATTTGTTGGTGCATATTTATGTGATGATCGGTTAC 492
Db      436 CAGTTCTTGGTTCTGGGTATCACAAACCGTTACTATTGATGCGATTGTTCATGTCGGATAC 495

QY      493 GCCACCCCTTCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGGCGCTG 552
Db      496 ACGGCACTGGCTGCGCAGCTAGGTGCGTATATTCGTTACCTAACATAATGACCAGAATG 555

QY      553 AATAAGATTTTGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTAGCATCGGCGA 607
Db      556 AATAAACTGTTGGTTTCGATGTTTATGGGCTGCGGAATGCTGTTGGCTACCGCCA 610
```

```
RESULT 13
US-10-282-122A-30630
; Sequence 30630, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30630
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30630

Query Match      18.3%; Score 112.4; DB 12; Length 630;
Best Local Similarity 53.5%; Pred. No. 1.2e-28;
Matches 324; Conservative 0; Mismatches 276; Indels 6; Gaps 4;

QY 16 TGGTTGCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCAGGCTCTGGTGCAATC 75
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 76 AACACTATGACCACCTCGCTCAACACAGGTTATCCGGCCGGTGGC--GTCTATTGCTGGG 133
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 134 CTTCAGACCGGACTGGCGATTTCATATTGTGCTGGTTGGCG-TGGGGTTGGACGCTATT 192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 136 CTGCAGATCGGCCTGGCCCTGCAGATCGCCATTGTGCGCGAGGCTACTGGAACCGCTGGGC 195
QY 193 TCCCGCTCAGTGATTGGCTTTGAAGTGTGAAGTGGGCAGGCGCGCTTACTTGTATTGG 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 196 GCGACCTCGGCACCTGGCTTTCAGCCTGATCAAGTGGTTGCGCGCGGCTGCTGTTGCTG 255
QY 253 CTGGGAATCCAGCAGTGGC--GCGCGCTGGTGCATTTGCAATTGACCTTAATCGCTGGCCTCTA 310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 311 CTCAATCG-CGTGACATTTGTTCCAGCGCGCAGTCTTTTGTGAATCTCACCATAATCCCAA 369
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 CTGGGCGGACCGTTGACCCCTGGTGTGCTGGTGTCTCTGGTCAACGCCAGCAATCCCAAG 375
QY 370 AGTATTGTGTTTCTGGCGGCGCTATTTCGGCAATTTCATCATGCCGCAACAGCCGCACTG 429
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 430 ATGCAGTATATCGTGTGCGCGTCAACCGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCG 549
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 490 TACGCCACCCCTTGCTCAACGGAATGCTCTATGGATTAAAGGACCAAGCAGATGAAGCG 549
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 550 CTGAATAAGATTTCGGCTCGTTGTTTATGCTGGTGGAGCGCTGTTAGCATCGGCGAGG 609
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 CATGCG 615
Db |||||

RESULT 14
US-10-282-122A-33154
; Sequence 33154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33154
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33154

Query Match      16.9%; Score 104; DB 12; Length 630;
Best Local Similarity 54.3%; Pred. No. 1.1e-25;
Matches 278; Conservative 0; Mismatches 225; Indels 9; Gaps 3;

QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACCCACCGGTATCCGGCCGGTGGC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GCGCGGGTGCCATCGCCTCGATGTCCTCGGCGCTGCAATACGGTTTCTCGCGGGCTAC 120
QY 121 GTCTATTGCT--GGGCTTCAGACCGGACTGGCGATTTCATATTGTGCTGTTGGCGTGGG 178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TGAATGCTTGGGCTGCAGATAGCGCTGGTGGCGCAGATCGCGATTGTCGGCGCGGG 180
QY 179 -TTGGAGCGCTATTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGCAGGCGCG 237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 CTTGGCGGGTATTGGCAGCTTCCGAGATGGCCTTCACGCTGATCAAGTGTTCGGCGTG 240
QY 238 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCG-----CCGCTGGTGCAATTGAC 291
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GCGTATCTGGTCTATCTGGGCATCAACAGTGGCGCGCATCGCCACCGACCTTGCAGC 300
QY 292 CTTAAATCGCTGGCCTCTACTCAATCGCGTGCAGATTGTTCCAGCGCGCAGTTTTTGTG 351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GAATCCGCGGTGCGGCGGTGGGCAAGCCGATGACGCTGGTGTTCGCGCGGTTTTCTGGTC 360
QY 352 AATCTACCAATCCCAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTCAATCATG 411
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 AACATCAGCAACCCCAAGCGCTGATCTTCAATCTGGCGATCTGCGCGAGTTCATCGAG 420
QY 412 CCGCAACAGCCGCAACTGATGCAGTATATCGTGTGCGGCTACCACTATTGTGTCGAT 471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 CCGACTGCTCCGCTGTTTCATGTCAGTACGTGATCATCGCCGCGACCATGGTCTGTCGAT 480
QY 472 ATTATTGTGATGATCGGTTACGCCACCTTGC 503
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 CTGATCGTCATGGCGGTTACACAGGCGTGGC 512
```

RESULT 15

US-10-282-122A-31699  
; Sequence 31699, Application US/10282122A  
; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 31699

; LENGTH: 633

; TYPE: DNA

; ORGANISM: Pseudomonas putida

US-10-282-122A-31699

Query Match 14.0%; Score 86.4; DB 12; Length 633;  
Best Local Similarity 49.5%; Pred. No. 1.9e-19;  
Matches 309; Conservative 0; Mismatches 306; Indels 9; Gaps 3;

QY	1	ATGACCTTAGAATGGTGGTTGCCTACCTGCTGACATCGATCAATTTAAACGCTGTCGCCA	60
Db	4	ATGTCGATGGAAGTATGGTAGGCTTTTGGCCCTGCTGGTGGATCAGCCTTTCACCC	63
QY	61	GGCTCTGGTGAATCAACACTATGACACCTCGTCAACACCGGTTATCCGGCCGGTGGC	120
Db	64	GGTCCGGGGCAATTGCCTCGATGTCCAGCGGCTGCAATACGGCTTCTGGCGTGGTTAC	123
QY	121	GTCTATTGC--TGGGCTTCAGACCGGACTGGCGATTATATTTGCTGGTGGCG-TGGG	177
Db	124	TGGAACGCCCTGGGCGCTGAGCTGGGCGCTGATCATGAGATCGCCATCATCGCCGCCGGC	183
QY	178	GTTGGGACGCTATTTCCCGCTCAGTGAATGGTTTGAAGTGTGAAGTGGGAGGCGCG	237
Db	184	GTCGGTGCCTCCTGGCTGGCTCGGCCAGCGGCTTCCAGGTCAATCAAGTTGGCGTC	243
QY	238	GCTTACTTGATTGGCTGGGAATCCAGAGTGGCGGCCGCT-----GGTGAATTGAC	291

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:07:39 ; Search time 1953.09 Seconds  
(without alignments)  
9403.168 Million cell updates/sec

Title: US-09-847-392-1\_COPY\_557\_1171

Perfect score: 615

Sequence: 1 atgacaccttagaatgtgtgtt.....tagcatcgcgagcgcattgog 615

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	127	20.7	473	28	AQ859250 nbeb0011L
2	85.4	13.9	1095	28	BZ548406 pacs1-50
3	78.6	12.8	769	28	AQ858095 nbeb0011L
C 4	55.8	9.1	1199	28	BZ566001 pacs2-164

C	5	53	8.6	231	14	CB042356	4007198 B
C	6	41	6.7	849	13	BX462111	BX462111
C	7	40.6	6.6	1039	29	CNS016GW	AL106730 Drosophil
C	8	39.8	6.5	990	29	AY401573	Mus muscu
C	9	39.4	6.4	712	13	BX416727	BX416727
C	10	39.2	6.4	1101	29	CNS0021D	AL097099 Drosophil
C	11	38.8	6.3	685	13	BY724027	BY724027
C	12	38.6	6.3	642	9	AL628057	AL628057
C	13	38	6.2	510	12	BG664436	BG664436 DRABDH02
C	14	37.6	6.1	475	28	BH209707	BH209707 Sml-41P23
C	15	37.6	6.1	798	10	BE559545	BE559545 GM700010A
C	16	37.2	6.0	885	13	BX425603	BX425603
C	17	37.2	6.0	1159	29	CNS015XR	AL106041 Drosophil
C	18	37	6.0	264	10	AW347581	AW347581 31605 MAR
C	19	37	6.0	308	10	AW266920	AW266920 0196-5 Le
C	20	37	6.0	437	12	BG688576	BG688576 336252 BA
C	21	37	6.0	450	10	AW482031	AW482031 40127 MAR
C	22	37	6.0	495	10	BF606721	BF606721 273837 MA
C	23	37	6.0	516	10	BE663253	BE663253 146023 MA
C	24	37	6.0	522	12	BM087647	BM087647 500452 MA
C	25	37	6.0	539	12	BM253190	BM253190 512572 MA
C	26	37	6.0	573	10	BE665319	BE665319 153519 MA
C	27	37	6.0	573	12	BI773985	BI773985 465757 MA
C	28	37	6.0	585	12	BI774021	BI774021 465798 MA
C	29	37	6.0	599	14	CB466945	CB466945 732659 MA
C	30	37	6.0	603	9	AV596505	AV596505 AV596505
C	31	37	6.0	629	14	CB462760	CB462760 723091 MA
C	32	37	6.0	1201	13	BX444391	BX444391
C	33	36.8	6.0	1201	13	BX385008	BX385008
C	34	36.4	5.9	1661	12	EM473385	EM473385 AGENCOURT
C	35	36	5.9	348	29	CE781017	CE781017 tigr-gss-
C	36	36	5.9	453	13	BY272658	BY272658
C	37	35.8	5.8	997	29	CNS006DN	AL065132 Drosophil
C	38	35.6	5.8	474	9	AL518216	AL518216
C	39	35.4	5.8	154	9	AV847594	AV847594
C	40	35.4	5.8	439	10	BE346900	BE346900 SP32e09.Y
C	41	35.4	5.8	547	9	AI965398	AI965398 sc71b10.Y
C	42	35.4	5.8	668	14	CF179229	CF179229 813328 MA
C	43	35.2	5.7	304	9	AL643781	AL643781
C	44	35.2	5.7	1201	13	BX403316	BX403316
C	45	35	5.7	931	13	BQ922252	BQ922252 AGENCOURT

ALIGNMENTS

RESULT 1

AQ859250/c

LOCUS

DEFINITION

AQ859250 473 bp DNA linear GSS 03-NOV-1999  
nbeb0011L14r CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica  
cultivar-group) genomic clone nbeb0011L14r, genomic survey  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

SUMMARY

AQ859250  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 473)  
Wing,R.A. and Dean,R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: GGAAACAGCTATGACCATG  
Class: BAC ends

High quality sequence start: 41  
High quality sequence stop: 415.  
Location/Qualifiers  
1..473

FEATURES

source

/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbeb0011114f"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC Library (EcoRI)"  
/notes="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;  
Rice is the most important food crop in the world. Half of  
the world population, especially those inhabiting highly  
populated areas of the humid tropics and subtropics, rely  
on rice as their primary source of carbohydrate.  
Monocotyledonous rice is a diploid plant (2n=24) with a  
haploid genome equivalent of 431 Mbp (Arumuganathan and  
Earle, 1991). The relatively small genome of rice, three  
times larger than that of Arabidopsis, makes it suitable  
for genomic studies. In order to facilitate positional  
cloning, physical mapping and genome sequencing of rice,  
we have constructed a BAC library from Oryza sativa,  
Nipponbare variety using EcoRI as the cloning enzyme. The  
library contains 55,296 clones with an average insert size  
of 121 Kb providing approximately 15 haploid genome  
equivalents. The deep coverage allows the isolation a  
particular sequence with a probability of 99.9 %. Three  
high density filters, each containing 18,432 clones  
(doubly spotted), represent the whole library for colony  
screening and can be requested from the Clemson University  
BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 20.7%; Score 127; DB 28; Length 473;  
Best Local Similarity 92.4%; Pred. No. 6.5e-27;  
Matches 145; Conservative 0; Mismatches 10; Indels 2; Gaps 1;  
QY 1 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 60  
Db 229 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGACATCGATCATTTTAACGCTGTGCGCA 170  
QY 61 GGCTCTGGTGCATCAACACATATGACCACTCGCTCAACACCGGTTAT--CCGGCGCGTG 118  
Db 169 GGCTCTGGTGCATCAACACATATGACCACTCGCTCAACACCGGTTATCGCGCGCGATA 110  
QY 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTC 155  
Db 109 AAGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTC 73

RESULT 2  
BZ548406  
LOCUS BZ548406 1095 bp DNA linear GSS 17-DEC-2002  
DEFINITION pacsl-60\_1079.sl pacsl-60 Pseudomonas aeruginosa genomic clone  
pacsl-60\_1079, genomic survey sequence.

ACCESSION BZ548406  
VERSION BZ548406.1 GI:27151987  
KEYWORDS GSS.

SOURCE

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
TITLE Whole-Genome-sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol. (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center

University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 20622216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES

source

Location/Qualifiers  
1..1095  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pacsl-60\_1079"  
/clone\_lib="pacsl-60"  
/note="clinical isolate 1-60 whole genomic shotgun  
library."

ORIGIN

Query Match 13.9%; Score 85.4; DB 28; Length 1095;  
Best Local Similarity 54.1%; Pred. No. 3.7e-14;  
Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;  
QY 218 TGTGAAGTGGGAGCGCGCTTACTTGTATGGCTGGGAATCCAGAGTGGC--GCGC 275  
Db 41 TGATCAAGTGGTTCGGCGTGGCTACCTGGTGTACCTGGCGGTGGCCAGTGGCAGCGC 100  
QY 276 CGCTGGTGCAATTGACCTTAATCGCTGGCCTCTACTCAATCG--CGTCGACATTTGTTCC 334  
Db 101 CGCCACAGGCTTGAGCACCGATGGCGAACGGCCTCTGGGGCGACCGTTGACCCCTGTTGC 160  
QY 335 AGCGGCGAGTTTTTGTGAATCTCACCATAATCCAAAGTATTGTGTTTCTGGCGGCGCTAT 394  
Db 161 TCGTGGTTTTCTGTCAACGCCAGCAATCCCAAGGGGGTGATCTTCATGTGTCGGGTGC 220  
QY 395 TTCGCGCAATTCATCGCCGCAACAGCCGCACTGATGCACTATATCGTGTGCGGCGTCA 454  
Db 221 TGCGCGAGTTTCATCGACCCGCGACCGCGCTGCTGGCGCAATACCTGATCATGGCGGCA 280  
QY 455 CCACATTTGTGTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGTCTCAACGGAATTG 514  
Db 281 CCATGATCGTCGTCGACCTGATCGTCATGCGCGGCTACACCGGGTGGCTGCGCGGTGC 340  
QY 515 CTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGT 574  
Db 341 TACGCGTACTGCGTTCGCCGCGCAGCAGAGCTGGTGAACCGTACCTTCGCCAGCCTGT 400  
QY 575 TTATGCTGGTGGAGCGCTGTTAGCATCGCGGAGGCATGCG 615  
Db 401 TCGTCGGTCCGCGGGGCTGTGGCGACCGGTACGCCGAGCG 441

RESULT 3

AQ858095

LOCUS AQ858095

DEFINITION

AQ858095 769 bp DNA linear GSS 03-NOV-1999  
nbeb0011114f CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica  
cultivar-group) genomic clone nbeb0011114f, genomic survey  
sequence.

ACCESSION

AQ858095

VERSION

AQ858095.1 GI:6208552

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 769)

REFERENCE

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA



Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clmson.edu  
Seq primer: TAATACGACTCACTATAGG  
Class: BAC ends  
High quality sequence start: 42  
High quality sequence stop: 454.  
Location/Qualifiers

FEATURES

source  
1..769  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbe001114f"  
/tissue\_type="Leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC Library (EcoRI)"  
/note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;  
Rice is the most important food crop in the world. Half of  
the world population, especially those inhabiting highly  
populated areas of the humid tropics and subtropics, rely  
on rice as their primary source of carbohydrate.  
Monocotyledonous rice is a diploid plant (2n=24) with a  
haploid genome equivalent of 431 Mbp (Arumuganathan and  
Earle, 1991). The relatively small genome of rice, three  
times larger than that of Arabidopsis, makes it suitable  
for genomic studies. In order to facilitate positional  
cloning, physical mapping and genome sequencing of rice,  
we have constructed a BAC library from Oryza sativa,  
Nipponbare variety using EcoRI as the cloning enzyme. The  
library contains 55,296 clones with an average insert size  
of 121 Kb providing approximately 15 haploid genome  
equivalents. The deep coverage allows the isolation a  
particular sequence with a probability of 99.9 %. Three  
high density filters, each containing 18,432 clones  
(doubly spotted), represent the whole library for colony  
screening and can be requested from the Clemson University  
BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 12.8%; Score 78.6; DB 28; Length 769;  
Best Local Similarity 94.5%; Pred. No. 3.4e-12;  
Matches 103; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTTAAAGCTGTGCGCA 60  
Dbb 658 ATGACCTTAGAATGGTGG-TTGCTACCTGCTGACATCGATCATTTAAGC-CTGTGCGCA 715  
Qy 61 GGCTCTGGTGCAATCAACACTATGACCACTGCTGCTCAACACGGTTATC 109  
Dbb 716 GGCTCTGGTGCAATCAACACTATGACCACTGCTGCTCAACACGGGTATC 764

RESULT 4

BZ566001/c  
LOCUS  
DEFINITION  
pac2-164\_5858.x1 pac2-164 Pseudomonas aeruginosa genomic clone  
pac2-164\_5858, genomic survey sequence.  
BZ566001  
VERSION  
BZ566001.1 GI:27194705  
KEYWORDS  
SOURCE  
ORGANISM  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1199)

REFERENCE

AUTHORS

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond

TITLE

JOURNAL

COMMENT

Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers

FEATURES

source  
1..1199  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pac2-164\_5858"  
/clone\_lib="pac2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun  
library."

ORIGIN

Query Match 9.1%; Score 55.8; DB 28; Length 1199;  
Best Local Similarity 50.1%; Pred. No. 3.3e-05;  
Matches 195; Conservative 0; Mismatches 187; Indels 7; Gaps 2;

Qy 128 GCTGGCTTCAGACCGGACTGGCGATTTCATATTGTGCTGGTGGCGTGGGTTGGGACGC 187  
Dbb 649 GCTGGCTTCAGATCGGGCTGCCCTGACAGATATCCCATTTGTTCTGCCGGGTGGTACG 590  
Qy 188 TATTTTCC--CGCTCAGTGATTGGCTTTGAAGTTGAAGTGGCAGGCGCGCTTACTT 245  
Dbb 589 TTACTGGGACCGTCGGCACTGGCTTACAGCATGATCAAGTTGCGCGTGGCTATCAT 530  
Qy 246 GATTGGCTGGGAATCCAGCAGTGGC----GCGCGCTGGTGCATTTGACCTTAAATCG 300  
Dbb 529 GGTGTACCTGCTGGTGGCCAGTGGCAGGCGCGCCACAGACCTTGAGCACCGATGGCG 470  
Qy 301 CTGGCCTTACTCAATCGCGCTGCAGATTTGTTCCAGCGCGCAGTTTGTGAATCTCACC 360  
Dbb 469 ACCGGCCCTCTGGTGCACCGTTGACCCCTGGTGGTGGTGCACCTGGTCAGCACACC 410  
Qy 361 AATCCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCAATTTCATCATGCGCGCAACAG 420  
Dbb 409 AATCACAAGACGGTGTCTGTCATGTCGCGGTGGTCCCGAGTTCAGCGACCGCATCAT 350  
Qy 421 CCGCAACTGATGCAGTATATCGTGGCGCGTGCACCACTATTGTGGTGCATATTATTGTG 480  
Dbb 349 CCGTGTGGCGCAAGACCTGATCATGGCGCGGCCCATGATCGTGTGTCGTCGTCGTCGTC 290  
Qy 481 ATGATCGGTTACGCCACCTTGTCTCAACG 509  
Dbb 289 ATGTCCGAGTACACCGGGGTGGTGGCGG 261

RESULT 5

BZ042356/c

LOCUS

DEFINITION  
BZ042356 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1P12  
Unknown, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos indicus

Bos indicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 231)

REFERENCE

AUTHORS

Wood, D.L., Capuco, A.V., Brito, M.A.P., Martinez, M.L., Connor, E.E.,

Machado, M.A. and Coutinho, L.L.

Construction and Characterization of cDNA Libraries Generated from

Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos

indicus) Cattle

Unpublished (2002)

COMMENT

Contact: Adilson F. da Mota  
Gene Evaluation and Mapping Laboratory  
USDA, ARS, Animal and Natural Resources Institute  
Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048456  
Fax: 3015048414  
Email: amota@cnpg1.embrapa.br  
Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim\_alt '' -trim\_fasta. Vector identified  
by cross\_match using options -minmatch 12 -minscore 18  
PCR Primers  
FORWARD: GTTTTCCCAGTCACGACGTTG  
BACKWARD: TGAGCGGATAACAATTCACACAG  
Plate: 1 row: P column: 12  
Seq primer: GTTTTCCCAGTCACGACGTTG  
High quality sequence stop: 231.

## FEATURES

Source

**ORIGIN**

Match	8.6%;	Score 53;	DB 14;	Length 231;
Local Similarity	56.6%;	Pred. No. 8.7e-05;		
98; Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0;
322	CGACATTTGTTCCAGCGCGCAGTTTTTTGTGAATCTCACCAATCCAAAAGTATTGTGTTT	381		
201	CGCCGTCCGTTCCTGGTTGGATTGCTCACCAATCTGACGAACCCGAAGGGCATCTCTGTT	142		
382	CTGGCGSGCTATTTCCGCAATTTCATCATGCCGCAACAGCCGCAACTGATGCAGTATATC	441		
141	ATGGTGGCCGTGCTGCCGCGAGTTTCATCGATCCGGCCAAAGCCGTGCTTTCGCAACTCGGC	82		
442	GTGCTCGGCGTCACCACTATTGTGGTCGATATTATTGTGATGATCGGTTACGC	494		
81	ATCCTGGCGGCGACGATGGTGTTCGTGCACCTGATCGTCAATGCACGGCTATGC	29		

RESULT 6	
BX462111/c	
LOCUS	849 bp mRNA linear EST 22-MAY-2003
DEFINITION	BX462111 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA clone CS0DG001YP07 5-PRIME, mRNA sequence.
ACCESSION	BX462111
VERSION	BX462111.1 GI:31029391
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 849)
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9966.f For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DG001CH04QP1&cluster=9966.f](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DG001CH04QP1&cluster=9966.f). Contact :  
Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> InvitroGen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DG001CH04QP1.

## ORIGIN

```

Query Match      6.7%; Score 41; DB 13; Length 849;
Best Local Similarity 21.3%; Pred. No. 0.77;
Matches 53; Conservative 86; Mismatches 110; Indels 0; Gaps 0;

QY 105 TTATCCGGCGGTGGCGTCTATTGTGGGCTTCAGACCGGACTGGCGATTTCATATTGTC 164
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 TTGTGCTGCTTTTGCTGCTTTGCTTTAGCNTTTGCGGCTATTGCAGGCGGCATTTC 441

QY 165 TGGTTGGCGTGGGTTGGGACGCTATTITCCCGTCACTGATTCGTTTGAAGTGTGAA 224
      ||| : ||| : : : : : ||| : ||| : : : : : ||| : : : : :
Db 440 TTGCKTGCKKGCKGCKKGCKKKKGCKTGCGCKKGCTSTSTTSKKSKKSGSKGS 381

QY 225 GTGGGCAGGCGCGCTTACTTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTC 284
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 KKSSTSAAGSKTGSSKKAKKKSSKKKSAKSGSKKGSKKGSKKGSKKGSNS 321

QY 285 AATTGACCTTAATCGGTGGCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 344
      : : : : : : : : : : : : : : : : : : : : : : :
Db 320 TGSNSCNSKKSAAKTATSGSGSKKKGSKTASTTSKKKSSKSKKSTSKKKSKSKSKKA 261

QY 345 TTTTGTGAA 353
      : : : : :
Db 260 CKKGSKKKA 252

```

RESULT 7	
CNS016GW	CNS016GW
LOCUS	1039 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN16O11 of DrosEAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106730
VERSION	AL106730.1 GI:5623541
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1039) Genoscope. Direct Submission
AUTHORS	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a





Sano,H., Sasaki,D., Sato,X., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES source

Location/Qualifiers  
1. .685  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="A230077K02"  
/sex="male"  
/tissue type="hypothalamus"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, adult male hypothalamus"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATAATATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 6.3%; Score 38.8; DB 13; Length 685;  
Best Local Similarity 53.0%; Pred. No. 3.1;  
Matches 105; Conservative 0; Mismatches 92; Indels 1; Gaps 1;  
  
QY 24 CTACCTGCTGACATCGATCATTTTAAAGCTGTGCGGCGGTCTATTGCTGGGCTTCAGACCTAT 83  
DB 443 CTCATTGTAGCCGTCGACAGTACAACACTTTCCTAGCCTCTACTTC-CTCGCGGTGAT 501  
  
QY 84 GACCACCTCGCTCAACACCGGTTATCCGGCCGGTGGCGTCTATTGCTGGGCTTCAGACCG 143  
DB 502 GAGCGCCGACCGATACCTGTGGTTCTGGCCACAGCAGAGTCGCGCGGGTGTCCGGGCG 561  
  
QY 144 GACTGGCGATTTCATATTGTGCTGGTGGGCTGGGCTGGGACGCTATTTCCTCGCTCAGT 203  
DB 562 CACTTACGGTGCAGCGCGGCTGTGCTGCTGGCGGTGGGCGCTGTGACCTGGTGTGCTGCT 621  
  
QY 204 GATTGCGTTTGAAGTGT 221  
DB 622 GCTGCCCTTTGCGGTATT 639

RESULT 12  
AL628057/c  
LOCUS  
DEFINITION  
AL628057 XGC-gastrula Silurana tropicalis cDNA clone TGas022011 5', mRNA linear EST 19-NOV-2003  
mRNA sequence.  
ACCESSION  
AL628057  
VERSION  
AL628057.2 GI:38448162  
KEYWORDS  
EST.  
SOURCE  
Silurana tropicalis (western clawed frog)  
ORGANISM  
Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.  
1 (bases 1 to 642)  
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
Unpublished (2003)  
On Nov 2, 2001 this sequence version replaced gi:16597540.  
Contact: Huckle E  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli XL1-blue  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TGas022011.plkSP6  
Sequencing primer: SP6.

FEATURES

Location/Qualifiers  
1. .642  
/organism="Silurana tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TGas022011"  
/dev\_stage="gastrula (stages 10.5-12 mixed)"  
/lab\_host="Escherichia coli XL1-blue"  
/clone\_lib="XGC-gastrula"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."  
end."

ORIGIN

Query Match 6.3%; Score 38.6; DB 9; Length 642;  
Best Local Similarity 50.8%; Pred. No. 3.5;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
  
QY 295 AAATCGCTGGCCTCTACTCAATCGCGTGCACATTTGTTCCAGCGCGAGTTTGTGAAT 354  
DB 252 AAATAGCATGAATTGTTCCATTGCTTCAAGTTTGTTTCATGGCTCAGTTTGTCTGG 193  
  
QY 355 CTCACCAATCCCAAAAGTATGTGTTTCTGGCGGCTATTTCGCAATTCATCATGCG 414  
DB 192 TTCTTCTGGCACTTACTGATGATGATTCAGGGAACACGATACCTGATCCTCCTGTCT 133  
  
QY 415 CAACAGCCGCAACTGATGCAGTATATCGTGCTCGGCGTCAACCACTATTGTGTCGATAT 474  
DB 132 CAGGAGCAGAGAGTGGCAGCAGTAGCCGGGGAGCCGCTGCTATTGCTGGGATGGAGACT 73  
  
QY 475 A 475  
DB 72 A 72

RESULT 13  
BG664436  
LOCUS  
DEFINITION  
BG664436 Rat DRG Library Rattus norvegicus cDNA clone DRABDH02 5', mRNA linear EST 30-APR-2001  
DRABDH02







SOURCE ORGANISM  
 Glycine max (soybean)  
 Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 1 (bases 1 to 798)  
 Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,  
 Erpelting, J., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.  
 A Functional Genomics Program for Soybean (NSF 9872565)  
 Unpublished (1999)  
 Other ESTs: A1965398 corresponding to Gm-cl016-1844 (5')  
 Contact: Vodka, L.O., PI, A Functional Genomics Program for  
 Soybean (NSF 9872565)  
 Lewin, H. A., Director, Keck Center for Comparative and Functional  
 Genomics  
 University of Illinois  
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
 Tel: (217) 244-6147  
 Fax: (217) 333-4582  
 Email: l-vodka@uiuc.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134. For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or info@genome  
 systems.com web site: www.genomesystems.com  
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

[illegible]

1. .798  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/\_clone="Gm-r1070-3841"  
/\_clone\_lib="Gm-r1070"  
/note="The library Gm-r1070 is a sequence-driven, reracked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones form the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzels, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html>.  
Reracking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

	Query Match	6.1%;	Score 37.6;	DB 10;	Length 798;
	Best Local Similarity	50.4%;	Pred. No. 7.9;		
	Matches 67;	Conservative	0;	Mismatches 66;	Indels 0; Gaps 0;
QY	398	CGCAATTTCATGCGCGCAACAGCCGCACTGATGCAGTATATCGTGTCTCGGCGTCACCA	457		
Db	738	CNNNATTNNNNCGTTAAACACACCCNNNTCTNNNNCTGTTATTATCTCTCATATTATCA	679		
QY	458	CTATGTGGTTCGATATTATTGTGATGATCGGTTACGCCACCCCTTGCTCAACCGATTGCTC	517		
Db	678	CTATCGTTATCGTTATCGTTGTGTTATCATCCCTTCCATCACCGTTCCTCAAAACACAGAGTCT	619		

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:08:16 ; Search time 60 Seconds  
(without alignments)  
965.371 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MTEWVFAYLTSIILTSP.....KIFGSLFVLVGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	3 AAY79298	E. coli R
2	1061	100.0	205	3 AAY99597	E. coli L
3	830.5	78.3	206	6 ABU32233	Protein e
4	827.5	78.0	206	6 ABU28287	Protein e
5	818.5	77.1	206	6 ABU47858	Protein e
6	803.5	75.7	190	4 ABG29208	Novel hum
7	791	74.6	205	6 ABU44939	Protein e
8	698.5	65.8	206	6 ABU50100	Protein e
9	686	64.7	138	4 AAG98975	E. coli g
10	686	64.7	138	6 ABU14756	Protein e
11	459.5	43.3	205	6 ABU49040	Protein e
12	410	38.6	245	4 ABG25225	Novel hum
13	389	36.7	209	6 ABU38890	Protein e
14	368	34.7	220	6 ADA33616	Acinetoba
15	367.5	34.6	210	6 ABU41414	Protein e
16	354.5	33.4	211	6 ABU39959	Protein e
17	312.5	29.5	209	6 ABU23178	Protein e
18	291	27.4	181	6 ABU16949	Protein e
19	290.5	27.4	211	6 ABU19902	Protein e
20	260.5	24.6	171	4 ABG29211	Novel hum
21	227	21.4	346	4 ABG29210	Novel hum
22	223	21.0	210	6 ADA35521	Acinetoba
23	215.5	20.3	241	6 ADA35872	Acinetoba
24	198	18.7	208	6 ADA33824	Acinetoba
25	195.5	18.4	212	3 AAB01787	Escherich

26	170	16.0	240	6	ADA36250	Acinetoba
27	168.5	15.9	235	6	ADA36962	Acinetoba
28	156.5	14.8	201	6	ABM68999	Photorhab
29	155	14.6	153	2	AAR87527	Mel-linke
30	155	14.6	153	2	AAW73358	S. colwel
31	150	14.1	207	4	AAG92664	C glutami
32	147	13.9	249	6	ADA34526	Acinetoba
33	146.5	13.8	205	6	ABM67507	Photorhab
34	144	13.6	130	3	AAB40407	Human ORF
35	144	13.6	130	5	ABP00692	Human ORF
36	134.5	12.7	226	4	AAG89911	C glutami
37	133.5	12.6	210	4	AAG81807	S. epider
38	133.5	12.6	224	5	ABP39763	Staphyloc
39	133	12.5	222	6	ADA35184	Acinetoba
40	126.5	11.9	213	6	ABM67782	Photorhab
41	126.5	11.9	223	3	AAB01786	Escherich
42	126	11.9	195	3	AAB01788	Escherich
43	124.5	11.7	214	6	ADA35408	Acinetoba
44	123.5	11.6	191	4	ABG29209	Novel hum
45	123.5	11.6	229	3	AAB11637	A. vitis

ALIGNMENTS

RESULT 1  
AAY79298  
ID AAY79298 standard; protein; 205 AA.  
XX  
AC AAY79298;  
XX  
DT 18-JUL-2000 (first entry)  
XX  
DE E. coli RhtB protein imparting homoserine resistance.  
XX  
KW Homoserine resistance; RhtB protein; L-homoserine; L-alanine;  
KW L-isoleucine; L-valine; L-threonine.  
XX  
OS Escherichia coli.  
XX  
PN EP994190-A2.  
XX  
PD 19-APR-2000.  
PF 20-SEP-1999; 93EP-00118581.  
XX  
PR 13-OCT-1998; 98RU-00118425.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Livshits VA, Zakataeva NP, Aleoshin VV, Belareova AV;  
PI Tokhmakova IL;  
XX  
DR WPI; 2000-273530/24.  
DR N-PSDB; AAZ94405.  
XX  
PT Novel RhtB protein, useful for generation of L-homoserine resistance in  
PT Escherichia bacteria and large-scale production of e.g. L-homoserine and  
PT L-alanine.  
XX  
PS Claim 1; Page 11-12; 14pp; English.  
XX  
CC The present sequence is that of the novel Escherichia coli K-12 protein,  
CC RhtB, which participates in resistance to L-homoserine. Amplification of  
CC the rhtB gene (see AAZ94405) results in an improvement of the amino acid  
CC productivity of E. coli. The invention provides: a protein comprising the  
CC present sequence or including a deletion, substitution, insertion and/or  
CC addition of 1 or more amino acids and having the activity of making a  
CC bacterium resistant to L-homoserine; DNA encoding RhtB; a bacterium,  
CC especially of the genus Escherichia, in which L-homoserine resistance is  
CC enhanced by amplifying the copy number or increasing the expression rate  
CC of the rhtB DNA, the DNA being carried on a multicopy vector or on a  
CC transposon; and a method for producing an amino acid by cultivating the

CC bacterium in a culture medium to produce and accumulate the amino acid in  
CC the medium, from which it is recovered. The method is used for the  
CC production of L-homoserine, L-alanine, L-isoleucine, L-valine or L-  
CC threonine (all claimed)

XX  
SQ Sequence, 205 AA;  
Query Match 100.0%; Score 1061; DB 3; Length 205;  
Best Local Similarity 100.0%; Pred. No. 1.2e-112;  
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTLWWFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWGV 60  
Db  
QY 1 MTLWWFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWGV 60  
Db  
QY 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLT 120  
Db 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLT 120  
QY 121 NPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180  
Db 121 NPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180  
QY 181 MKALNKIFGSLFMLVGALLASARHA 205  
Db 181 MKALNKIFGSLFMLVGALLASARHA 205

RESULT 2  
AAY99597  
ID AAY99597 standard; protein; 205 AA.  
XX  
AC AAY99597;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE E. coli L-homoserine resistance protein, RhtB.  
XX  
KW L-homoserine resistance; L-homoserine synthesis; rhtB; L-threonine;  
KW L-valine; L-leucine.  
XX  
OS Escherichia coli.  
XX  
PN EP1013765-A1.  
XX  
PD 28-JUN-2000.  
XX  
PF 20-DEC-1999; 99EP-00125406.  
XX  
PR 23-DEC-1998; 98RU-00123511.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
PI Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;  
XX  
DR WPI; 2000-414602/36.  
DR N-PSDB; AAA48442.  
XX  
PT Novel Escherichia bacterium having enhanced L-threonine resistance due to  
PT enhanced RhtC protein activity, used to produce L-threonine, L-  
PT homoserine, L-valine and L-leucine.

PS Claim 2; Page 13-14; 24pp; English.  
XX  
XX The present sequence is the L-homoserine resistance protein, RhtB, from  
CC Escherichia coli. The coding sequence may be used to impart L-homoserine  
CC resistance on E. coli bacteria, which would be useful for producing a  
CC high yield of L-homoserine. L-homoserine resistance means that the  
CC bacteria will be able to grow on a minimal medium containing L-homoserine  
CC at a concentration at which the corresponding wild-type strain would not  
CC grow. Since the transformed bacteria can grow on the minimal medium, it  
CC can synthesize L-homoserine, which accumulates. The accumulated amino  
CC acids can then be removed from the culture medium. The bacterium of the

CC present invention may also be used to synthesise L-threonine, L-valine  
CC and L-leucine at increased levels  
XX  
SQ Sequence 205 AA;

Query Match 100.0%; Score 1061; DB 3; Length 205;  
Best Local Similarity 100.0%; Pred. No. 1.2e-112;  
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTLWWFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWGV 60  
Db  
QY 1 MTLWWFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWGV 60  
Db  
QY 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLT 120  
Db 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLT 120  
QY 121 NPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180  
Db 121 NPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180  
QY 181 MKALNKIFGSLFMLVGALLASARHA 205  
Db 181 MKALNKIFGSLFMLVGALLASARHA 205

RESULT 3  
ABU32233  
ID ABU32233 standard; protein; 206 AA.  
XX  
AC ABU32233;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #17760.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA36103.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 60157; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 206 AA;

Query Match 78.3%; Score 830.5; DB 6; Length 206;  
Best Local Similarity 81.2%; Pred. No. 2.6e-86;  
Matches 169; Conservative 13; Mismatches 21; Indels 5; Gaps 3;  
QY 1 MTLEWTFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTG--DSYCAGW 57  
DB 1 MTLEWTFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTG--DSYCAGW 58  
QY 58 RGVGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV 117  
DB 59 VGLGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV 118  
QY 118 NLTPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRIALWIKG 177  
DB 119 NLTPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRIALWIKG 178  
QY 178 PKQMKALNKIFGSLFMLVGLLASARHA 205  
DB 179 PKQMKALNKIFGSLFMLVGLLASARHA 206

RESULT 4  
ABU28287  
ID ABU28287 standard; protein; 206 AA.  
XX  
AC ABU28287;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #13814.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Enterobacter cloacae.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA32157.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 56211; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 206 AA;

Query Match 78.0%; Score 827.5; DB 6; Length 206;  
Best Local Similarity 79.1%; Pred. No. 5.7e-86;  
Matches 167; Conservative 15; Mismatches 18; Indels 11; Gaps 3;  
QY 1 MTLEWTFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGSDSYCAGW-- 58  
DB 1 MTLEWTFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGSDSYCAGW-- 55  
QY 59 ----GVGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORA 114  
DB 56 LVGIGLGLTFLSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORA 115  
QY 115 VFNLTNPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRIALW 174  
DB 116 VFNLTNPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRIALW 175  
QY 175 IKGPKQMKALNKIFGSLFMLVGLLASARHA 205  
DB 176 IKGPKQMKALNKIFGSLFMLVGLLASARHA 206

RESULT 5  
ABU47858  
ID ABU47858 standard; protein; 206 AA.



CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 190 AA;

Query Match 75.7%; Score 803.5; DB 4; Length 190;  
Best Local Similarity 88.9%; Pred. No. 2.8e-83;  
Matches 169; Conservative 3; Mismatches 13; Indels 5; Gaps 3;  
QY 19 SPGSGAINTMTTSLNHGYPAGGVYCWASDRTG---DSYCAGWRGVGTLSRSRVIAFEVLK 75  
Db 3 SPGSGAINTMTTSLNHGY-RGAVASIAGLQGLAIHIVLVG-VGLGTLFSRSRVIAFEVLK 60  
QY 76 WAGAAYLWLGIQOWRAAGAILDKSLASTQSRHLFORAVFVNLTNPKSIVFLAALFPQF 135  
Db 61 WAGAAYLWLGIQOWRAAGAILDKSLASTQSRHLFORAVFVNLTNPKSIVFLAALFPQF 120  
QY 136 IMPOQPQLMQYIVLGVTIVVDIIVMIGYATLAQRIALWIKGPKOMKALNKIFGSLFMLV 195  
Db 121 IMPOQPQLMQYIVLGVTIVVDIIVMIGYATLAQRIALWIKGPKOMKALNKIFGSLFMLV 180  
QY 196 GALLASARHA 205  
Db 181 GALLASAKHA 190

RESULT 7  
ABU44939  
ID ABU44939 standard; protein; 205 AA.  
XX  
AC ABU44939;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #30466.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Salmonella paratyphi.

XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA48809.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 72863; 1766pp; English.

XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 205 AA;

Query Match 74.6%; Score 791; DB 6; Length 205;  
Best Local Similarity 76.8%; Pred. No. 8.4e-82;  
Matches 162; Conservative 14; Mismatches 23; Indels 12; Gaps 4;  
QY 1 MTLWNFPAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGSDSYCAGWR-- 58  
Db 1 MTFEWNFPAYLLTSTLLSLSPGSGAINTMTTSLNHGY-RGAAASIAGLQTG----LGIHIV 55  
QY 59 ----GVGTLSRSVIAFEVLKWAGAAAYLIWLGIQOWRAAGAILDKSLASTQSRHLFORA 114  
Db 56 LVGVGLGTLSRSVIAFEVLKWAGAAAYLIWLGIQOWR-AGAILDLHTLAQTQSRGRLFKRA 114  
QY 115 VFVNLTNPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTIVVDIIVMIGYATLAQRIALW 174  
Db 115 IFVNLTNPKSIVFLAALFPQFIMPOQPQLAQYLILGVTTIVVDIMIVMTGYATLAXRTAAW 174  
QY 175 IKGPKQMKALNKIFGSLFMLVGALLASARHA 205  
Db 175 IKGPKQMKALNKAFGSLFMLVGALLASARHA 205

RESULT 8  
ABU50100  
ID ABU50100 standard; protein; 206 AA.  
XX  
AC ABU50100;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #35627.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Yersinia pestis.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.



PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA53970.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 78024; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 206 AA;  
Query Match 65.8%; Score 698.5; DB 6; Length 206;  
Best Local Similarity 63.3%; Pred. No. 3.1e-71;  
Matches 136; Conservative 33; Mismatches 23; Indels 23; Gaps 3;  
Qy 1 MTLWWEFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWAASDRGDSYCAGWR-- 58  
Db 1 MTLDWLTYLLTTLTSLSPGSGAINTMTSTAISHG--TRGVV-----ASIGGLQLG 49  
Qy 59 -----GVGTFRSRVIAFEVLKWAGAYLIWLGIQWRAGAILDKSLASTQSR 108  
Db 50 LAVHIVLVGVGLGALVQSLLAFEPILKWLGAAYLIWLGIQWRAGAILDKSLASTQSR 109  
Qy 109 HLFQRAVFNLTNPXSIVFLAALFPQFIMPQPOLMQYIVLGVTTIVDIIWMIGYATLA 168  
Db 110 KLFKRAVFNLTNPXSIVFLAALFPQFVLPQFPQVAQYLIIGLSTSVIVDIIWMIGYATLA 169  
Qy 169 QRIALWIKGPKQMKALNKIFGSLFMLVGALLASAR 203  
Db 170 TRIARWIKSPQOMKLLNRIFGGLFMLIGALLATAR 204

RESULT 9  
AAG98975  
ID AAG98975 standard; protein; 138 AA.  
XX  
XX AC AAG98975;  
XX  
XX 26-SEP-2001 (first entry)  
XX  
XX E. coli growth and proliferation related protein sequence SEQ ID NO:445.  
DE Escherichia coli; growth; proliferation; microbial; antimicrobial;  
XX bacterial infection; microorganism.  
KW Escherichia coli.  
XX  
OS WO200134810-A2.  
XX  
XX 17-MAY-2001.  
PD  
XX 09-NOV-2000; 2000WO-US030950.  
PF  
XX 09-NOV-1999; 99US-0164415P.  
PR (ELIT-) ELITRA PHARM INC.  
PA Forsyth RA, Ohlsen K, Zyskind J;  
XX  
PI WPI; 2001-335933/35.  
XX N-PSDB; AAH84646.  
DR  
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
PT for screening for homologous genes and for designing expression vectors.  
XX  
XX Claim 19; Page 502-503; 522pp; English.  
XX  
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli growth  
CC and proliferation related proteins given in AAG99078 and AAG98830 to  
CC AAG98999. (I) can be used as potential targets for the generation of new  
CC antimicrobial agents, and for identification of compounds which interact  
CC with the gene products of (I). In addition the expression of (I) and the  
CC purification of the proteins, the purified proteins can be used to  
CC generate reagents and screen small molecule libraries or other candidate  
CC compound libraries for compounds that can be further developed to yield  
CC novel antimicrobial compounds. In addition, nucleic acid probes  
CC complementary to (I) that are specific for particular species of  
CC microorganisms can be used to identify particular microorganism species  
CC in clinical specimens, therefore, providing a rapid and dependable method  
CC by which to identify the causative agents of a bacterial infection. Also,  
CC antibodies generated against proteins translated from mRNA transcribed  
CC from proliferation-required sequences can also be used to screen for  
CC specific microorganisms that produce such proteins in a species-specific  
CC manner. AAH84371 and AAH84670 represent sequencing primers used in the  
CC isolation of E. coli growth and proliferation related sequence, which are  
CC used in an example from the present invention  
XX  
SQ Sequence 138 AA;

Query Match 64.7%; Score 686; DB 4; Length 138;  
Best Local Similarity 99.3%; Pred. No. 4.9e-70;  
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 68 VIAFEVLKWAGAYLIWLGIQWRAGAILDKSLASTQSRRLHFQRAVFNLTNPXSIVF 127  
Db 1 MIAFEVLKWAGAYLIWLGIQWRAGAILDKSLASTQSRRLHFQRAVFNLTNPXSIVF 60  
Qy 128 LAALFPQFIMPQPOLMQYIVLGVTTIVDIIWMIGYATLAORIALWIKGPKQMKALNKI 187  
Db 61 LAALFPQFIMPQPOLMQYIVLGVTTIVDIIWMIGYATLAORIALWIKGPKQMKALNKI 120  
Qy 188 FGSLEFMLVGALLASARHA 205  
Db 121 FGSLEFMLVGALLASARHA 138

```

RESULT 10
ABU14756
ID ABU14756 standard; protein; 138 AA.
XX
AC ABU14756;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #283.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Escherichia coli.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA18626.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 42680; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 138 AA;
XX
Query Match 64.7%; Score 686; DB 6; Length 138;
Best Local Similarity 99.3%; Pred. No. 4.9e-70;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 68 VIAFEVLKWAGAAVLIWLGIIQQWRAAGAILDKSLASTQSRRLHFQRAVFNLTNPKSIVF 127
DB 1 MIAFEVLKWAGAAVLIWLGIIQQWRAAGAILDKSLASTQSRRLHFQRAVFNLTNPKSIVF 60
QY 128 LAALFPQFIMPQQPQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKGPKQMKALNKI 187
DB 61 LAALFPQFIMPQQPQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKGPKQMKALNKI 120
QY 188 FGSLEFVLGALLASARHA 205
DB 121 FGSLEFVLGALLASARHA 138
RESULT 11
ABU49040
ID ABU49040 standard; protein; 205 AA.
XX
AC ABU49040;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #34567.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Vibrio cholerae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA52910.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 76964; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

```



PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA42760.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 66814; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 209 AA;  
SQ  
Query Match 36.7%; Score 389; DB 6; Length 209;  
Best Local Similarity 36.4%; Pred. No. 7.4e-36;  
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;  
QY 1 MTLEWFAVLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58  
Db 1 MLVSTWFAFFLACWAISLSPGAGAIASMSGCGLYGFARG-----YWNAL 44  
QY 59 -----GVGTLFSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAILDKLSLAS 103  
Db 45 GLQIGLALQIAIVAAGVGALLATSALAFSLIKWFGVAYLVILAVRQWQAP-----PQALS 99  
QY 104 TQSRRLHFQ-----RAVFVNLTNPKSIVFLAALFPQFIMPPQQLMQYIVLGVTTIVVD 157  
Db 100 TDGERELGRPLTLVLRGFLVNASNPKAVIFMLAVLPQFIDPHQPLLAQYLVINGGTMIVVD 159  
QY 158 IIVMIGYATLAQRALNWKPKQMKALNKIFGSLFMLVGLLASARHA 205  
Db 160 LIVMAGYTGLAARVLRVLRSPRQQLVNRFTFASLVFGAAGLLATVRR 207

RESULT 14  
ADA33616  
ID ADA33616 standard; protein; 220 AA.  
XX  
AC ADA33616;  
XX  
DT 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #777.  
DE  
XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
KW plant biocontrol agent.  
XX Acinetobacter baumannii.  
OS  
XX US6562958-B1.  
PN  
XX 13-MAY-2003.  
PD  
XX 04-JUN-1999; 99US-00328352.  
PF  
XX 09-JUN-1998; 98US-0089701P.  
PR  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX Breton G, Bush D;  
PI  
XX WPI; 2003-576092/54.  
DR  
XX N-PSDB; ADA29490.  
DR  
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
XX for diagnosing a bacterial disease, as components of antibacterial  
XX PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
XX PT plants.  
XX  
XX Example; SEQ ID NO 4903; 328pp; English.  
PS  
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents the amino acid sequence of an A.  
CC baumannii protein.  
XX  
SQ Sequence 220 AA;  
Query Match 34.7%; Score 368; DB 6; Length 220;  
Best Local Similarity 35.3%; Pred. No. 2e-33;  
Matches 77; Conservative 46; Mismatches 71; Indels 24; Gaps 3;  
QY 1 MTLEWFAVLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58  
Db 13 MSLQVWFAYMLACWVISISPGAGAIASMSGSLNYGF-----RHGYWNAIGLQIA 61  
QY 59 -----GVGTLFSRSVIAFEVLKWAGAAAYLIWLGIQWRA-AGAILDKLSLASTQSR 107  
Db 62 LLIQIMIVAAGVGVLFAATPLAFAQAVKWFVAYLLYLAYLQWTAPVKDIEIQHEKKDSV 121  
QY 108 RHLFQRAVFVNLTNPKSIVFLAALFPQFIMPPQQLMQYIVLGVTTIVVDIIVMIGYATL 167  
Db 122 SALLFNGFVNISNPKAIVFLLAVLPQFLDLSKQWQIYLIIMAAATMTVIDLIVNAGYTGL 181  
QY 168 AQRIALWIKGPKQMKALNKIFGSLFMLVGLLASARHA 205  
Db 182 ASKVLRLRSRSPKQKYLNRGFAVMFSCAALLLSTVHQA 219  
RESULT 15  
ABU41414  
ID ABU41414 standard; protein; 210 AA.  
XX  
AC ABU41414;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #26941.  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:13:32 ; Search time 22 Seconds  
(without alignments)  
481.060 Million cell updates/sec

Title: US-09-847-392-2  
Perfect score: 1061  
Sequence: 1 MILEWFWAYLLTSIILTLSP.....KIFGSLFMLVGALLASARHA 205

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	4	US-09-396-357-2
2	832.5	78.5	214	4	US-09-489-039A-8066
3	686	64.7	138	4	US-09-711-164-445
4	389	36.7	277	4	US-09-252-991A-23018
5	368	34.7	220	4	US-09-328-352-4903
6	228	21.5	250	4	US-09-543-681A-4487
7	223	21.0	210	4	US-09-328-352-6808
8	215.5	20.3	241	4	US-09-328-352-7159
9	200.5	18.9	214	4	US-09-489-039A-9793
10	198	18.7	208	4	US-09-252-991A-26644
11	193	18.2	219	4	US-09-489-039A-7731
12	193	18.2	222	4	US-09-252-991A-26644
13	179.5	16.9	237	4	US-09-252-991A-27175
14	175.5	16.5	211	4	US-09-252-991A-20747
15	170	16.0	240	4	US-09-328-352-7537
16	168.5	15.9	235	4	US-09-328-352-8249
17	164.5	15.5	228	4	US-09-543-681A-4854
18	163.5	15.4	212	4	US-09-543-681A-4767
19	156	14.7	216	4	US-09-252-991A-18666
20	155	14.6	153	2	US-08-476-254-7
21	155	14.6	153	6	5474933-4
22	148	13.9	226	4	US-09-252-991A-31610
23	147.5	13.9	211	4	US-09-489-039A-9608
24	147	13.9	249	4	US-09-328-352-5813
25	144	13.6	217	4	US-09-489-039A-8076
26	140	13.2	260	4	US-09-252-991A-25992
27	138.5	13.1	220	4	US-09-543-681A-4545

28	133.5	12.6	224	4	US-09-134-001C-4608	Sequence 4608, Ap
29	133	12.5	222	4	US-09-328-352-6471	Sequence 6471, Ap
30	130	12.3	103	4	US-09-543-681A-7951	Sequence 7951, Ap
31	124.5	11.7	214	4	US-09-328-352-6695	Sequence 6695, Ap
32	123.5	11.6	206	4	US-09-252-991A-18746	Sequence 18746, A
33	123.5	11.6	211	4	US-09-489-039A-10442	Sequence 10442, A
34	119	11.2	271	4	US-09-252-991A-32455	Sequence 32455, A
35	116.5	11.0	206	4	US-09-328-352-4784	Sequence 4784, Ap
36	116	10.9	213	4	US-09-252-991A-16947	Sequence 16947, A
37	113.5	10.7	269	4	US-09-489-039A-7320	Sequence 7320, Ap
38	113	10.7	209	4	US-09-328-352-8225	Sequence 8225, Ap
39	112.5	10.6	200	4	US-09-328-352-4773	Sequence 4773, Ap
40	109	10.3	249	4	US-09-252-991A-25717	Sequence 25717, A
41	107.5	10.1	221	4	US-09-328-352-6659	Sequence 6659, Ap
42	104	9.8	210	4	US-09-328-352-4775	Sequence 4775, Ap
43	102	9.6	213	4	US-09-489-039A-11845	Sequence 11845, A
44	100	9.4	200	4	US-09-328-352-4827	Sequence 4827, Ap
45	98	9.2	214	4	US-09-328-352-6025	Sequence 6025, Ap

ALIGNMENTS

RESULT 1  
US-09-396-357-2  
; Sequence 2, Application US/09396357  
; Patent No. 6303348  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERES ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/396,357  
; CURRENT FILING DATE: 1999-09-15  
; EARLIER APPLICATION NUMBER: RU98118425  
; EARLIER FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-396-357-2

Query Match		100.0%;	Score 1061;	DB 4;	Length 205;
Best Local Similarity		100.0%;	Pred. No. 1,le-116;		
Matches 205;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MILEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWRGV	60		
Db	1	MILEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWRGV	60		
Qy	61	GTLPFSRSVIAFEVLKWAGAYLIWLGIQWRAAGAIIDLKSLASTQSRRLHFQRAVFNLT	120		
Db	61	GTLPFSRSVIAFEVLKWAGAYLIWLGIQWRAAGAIIDLKSLASTQSRRLHFQRAVFNLT	120		
Qy	121	NPXSIVFLAALFPQFIMPQQLMQYIVLGVTTIVVDIWMIGYATLAQRIALWIKGPQK	180		
Db	121	NPXSIVFLAALFPQFIMPQQLMQYIVLGVTTIVVDIWMIGYATLAQRIALWIKGPQK	180		
Qy	181	MKALNKIFGSLFMLVGALLASARHA	205		
Db	181	MKALNKIFGSLFMLVGALLASARHA	205		

RESULT 2  
US-09-489-039A-8066  
; Sequence 8066, Application US/09489039A



```

; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8066
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8066

Query Match      78.5%; Score 832.5; DB 4; Length 214;
Best Local Similarity 81.7%; Pred. No. 8.9e-90;
Matches 170; Conservative 12; Mismatches 21; Indels 5; Gaps 3;

QY 1 MTEWTFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWAASDRIG---DSYCAGW 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 9 MTIEWWFAYLLTSIIITLSPGSGAINTMTTSLNHGY-RGAASIAAGLQGLAIHIVLG- 66
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 58 RGVGTIFSRSVIAFEVLKWAGAAVLIIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFV 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 67 VGLGTIFSRSVLAFEVLKWAGAAVLIIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFV 126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 118 NLTNPKSIVFLAALFPQFIMPPQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWTKG 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 127 NLTNPKSIVFLAALFPQFIMPPQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWTKG 186
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 178 PKQMKALNKIFGSLFMLVGALLASARHA 205
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 PKQMKALNKVFGSLFMLVGALLASARHA 214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
US-09-711-164-445
; Sequence 445, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 445
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-445

Query Match      64.7%; Score 686; DB 4; Length 138;
Best Local Similarity 99.3%; Pred. No. 8.1e-73;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAFEVLKWAGAAVLIIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFVNLTPKSIIVF 127
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MIAFEVLKWAGAAVLIIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFVNLTPKSIIVF 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 128 LAALFPQFIMPPQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQMKALNKI 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LAALFPQFIMPPQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQMKALNKI 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 188 FGSFLMLVGALLASARHA 205
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 121 FGSFLMLVGALLASARHA 138
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
US-09-252-991A-23018
; Sequence 23018, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23018
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23018

Query Match      36.7%; Score 389; DB 4; Length 277;
Best Local Similarity 36.4%; Pred. No. 1.8e-37;
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;

QY 1 MTEWTFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWAASDRIGTSDSYCAGWR-- 58
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 69 MLVSTWFAFFLACWAIISLSPGAGIASMSCGLQYGFARG-----YWNAL 112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 59 -----GVGTLFPSRSVIAPEVLKWAGAAVLIIWLGIQOWRAAGAILDKSLAS 103
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 113 GLQIGLALQIAIVAAGVGALLATSALAFSLIKMFGVAYLVYLAVRQWQAP-----POALS 167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 104 TQSRRLHFQ-----RAVFVNLTPKSIIVFLAALFPQFIMPPQPOLMQYIVLGVTTIVVD 157
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 168 TDGERPLGRPLTLVLRGFLVNASNPKAVIFMLAVLPQFIDPHQPLLAQYLLIMGTMIVVD 227
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 158 IIVMIGYATLAQRIALWIKGPKQMKALNKIFGSLFMLVGALLASARHA 205
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 228 LIVMAGYTGLAARVLRLSPRQQLVNRFTFASLFFVGAAGLLATVRA 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-09-328-352-4903
; Sequence 4903, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4903
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4903

Query Match      34.7%; Score 368; DB 4; Length 220;
Best Local Similarity 35.3%; Pred. No. 3.8e-35;
Matches 77; Conservative 46; Mismatches 71; Indels 24; Gaps 3;

QY 1 MTEWTFAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWAASDRIGTSDSYCAGWR-- 58
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 MSLOVWFAYMLACWVISISPGAGIASMSGLNYGF-----RHGYWNAIGLQIA 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```





Db 196 RAQRLOQWGFGLLIGFVRLALLR 220

RESULT 13

US-09-252-991A-27175

; Sequence 27175, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27175

; LENGTH: 237

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27175

Query Match 16.9%; Score 179.5; DB 4; Length 237;

Best Local Similarity 27.9%; Pred. No. 6.2e-13;

Matches 60; Conservative 43; Mismatches 87; Indels 25; Gaps 9;

QY 4 EWWFAYLTSIIILTLSPGSGAINTMTTSLNHGYPAG-----GVYCWASDRTGDS--YCAG 56

Db 29 DFW-TYVLGVVFVILLPGPNSFLVLTSAQRGVATGYRAACGVF-----LGDAVLMLLS 81

QY 57 WRGVGTLPFSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDLKSLASTQSR-----RHL 110

Db 82 ALGVASLLKAEPMFLIGLKYLGAAAYLFYLGVMRLGAWR-KLRNPEATAAQAEKVDVHQP 140

QY 111 FQRAVFNLTNPKSIVFLAALFPQFIMP--QOPQLMQYIVLGVTTIVVDIIVMIGYATLA 168

Db 141 FRKALLLSLSNPKAILFFISFFIQFVDPGYAYPGL-SFLVLAVILELVSAALYLSFLIFTG 199

QY 169 QRIALWIKGPKQKMA-LNKIFGSLFMLVGLLASA 202

Db 200 VRLAAWERRRQRLAAGATSGVGALFVGFVGLATA 234

RESULT 14

US-09-252-991A-20747

; Sequence 20747, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20747

; LENGTH: 211

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20747

Query Match 16.5%; Score 175.5; DB 4; Length 211;

Best Local Similarity 26.0%; Pred. No. 1.6e-12;

Matches 56; Conservative 34; Mismatches 98; Indels 27; Gaps 6;

QY 6 WFAYLTSIIILTLSPGSGAINTMTTSLNHGYP-----AGGVYCWASDRTGDSYCAGW----- 57

Db 9 WALFVPACFALNLAGPNNLLSNNAAARHGFFATASLAGGRLLA-----FAGMLALA 60

QY 58 -RGVGTLPFSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDLKSLASTQSR-----HLF 111

Db 61 ASGLALVLTSAWLFLAIKVLGAAVLLNLAVQLWRT----DAQPLVNEASPARSLWRLG 116

QY 112 QRAVFNLTNPKSIVFLAALFPQFIMPQOPQLMQYIVLGVTTIVVDIIVMIGYATLAQRI 171

Db 117 ROEFLVAGNPKAILIFTAFLPQFVDPGQPLGAFQAGLGAAPFLLEWLAIALYSYAGLHL 176

QY 172 ALWIKGPKQKMA-LNKIFGSLFMLVG-ALLASARHA 205

Db 177 GRLLAGQRRARLEFNRGCAALLGSAGLGLLSRRPA 211

RESULT 15

US-09-328-352-7537

; Sequence 7537, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 7537

; LENGTH: 240

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-7537

Query Match 16.0%; Score 170; DB 4; Length 240;

Best Local Similarity 28.0%; Pred. No. 8.3e-12;

Matches 59; Conservative 36; Mismatches 78; Indels 38; Gaps 9;

QY 9 YLLTSIIILTLSPGSGAINTMTTSLNHGYPAG-----GVYCWASDRTGD---SYCAGWRGV 60

Db 34 YIIGTFLIVLLPGPNSLYVMSIASRYGIKTGYMGALGIF-----TGDLLMLCT-VLGA 86

QY 61 GTLFSRSVIAFEVLKWAGAAAYLIWLG-----IQQWR-----AAGAIIDLKSLASTQSR 108

Db 87 ASLLKAFEPWFIVLKLVLGALYLSYLGFKLLQGSIQRWKLRNQPOAEMADLPAL----DKV 142

QY 109 HLFQRAVFNLTNPKSIVFLAALFPQFIMPQOP-QLMQYIVLGVTTIVVDIIVMIGYATL 167

Db 143 HPYKTALISLLNPKAILFFLSFFQFVEPDYAYPALSFLLAV-----ILQIISFSYL 196

QY 168 AQRIALWIKGPKQKMA-LNKIFGSLFMLVGL 198

Db 197 TALIFSGIKLSTFFKQNHKIAASGIFLVGIL 227

Search completed: March 24, 2004, 21:17:17

Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:16:17 ; Search time 41 Seconds  
(without alignments)  
1294.777 Million cell updates/sec

Title: US-09-847-392-2  
Perfect score: 1061  
Sequence: 1 MLEWFWFAYLLTSIILTLSP.....KIFGSLFMLVGALLASARHA 205

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	9	US-09-927-395-2
2	1061	100.0	205	9	US-09-847-392-2
3	830.5	78.3	206	12	US-10-282-122A-60157
4	827.5	78.0	206	12	US-10-282-122A-56211
5	818.5	77.1	206	12	US-10-282-122A-75782
6	791	74.6	205	12	US-10-282-122A-72863
7	698.5	65.8	206	12	US-10-282-122A-78024
8	686	64.7	138	12	US-10-282-122A-42680
9	686	64.7	138	14	US-10-287-274-445
10	459.5	43.3	205	12	US-10-282-122A-76964
11	389	36.7	209	12	US-10-282-122A-66814
12	367.5	34.6	210	12	US-10-282-122A-69338
13	354.5	33.4	211	12	US-10-282-122A-67883
14	312.5	29.5	209	12	US-10-282-122A-51102
15	291	27.4	181	12	US-10-282-122A-44873

16	290.5	27.4	211	12	US-10-282-122A-47826	Sequence 47826, A
17	180.5	17.0	217	14	US-10-156-761-10918	Sequence 10918, A
18	163.5	15.4	212	12	US-10-374-903A-6	Sequence 6, Appli
19	150	14.1	207	9	US-09-738-626-6418	Sequence 6418, Ap
20	147	13.9	224	14	US-10-156-761-7796	Sequence 7796, Ap
21	139.5	13.1	205	14	US-10-156-761-11297	Sequence 11297, A
22	134.5	12.7	226	9	US-09-738-626-3665	Sequence 3665, Ap
23	126	11.9	195	12	US-10-620-487-2	Sequence 2, Appli
24	105.5	9.9	223	9	US-09-738-626-6070	Sequence 6070, Ap
25	105.5	9.9	223	10	US-09-746-660A-14	Sequence 14, Appl
26	94	8.9	456	12	US-10-282-122A-75400	Sequence 75400, A
27	94	8.9	457	12	US-10-282-122A-43084	Sequence 43084, A
28	94	8.9	457	15	US-10-369-493-729	Sequence 729, App
29	94	8.9	463	9	US-09-815-242-13918	Sequence 13918, A
30	93	8.8	448	12	US-10-282-122A-72853	Sequence 72853, A
31	92.5	8.7	363	12	US-10-282-122A-46325	Sequence 46325, A
32	92	8.7	456	12	US-10-282-122A-74836	Sequence 74836, A
33	91	8.6	457	12	US-10-282-122A-60097	Sequence 60097, A
34	91	8.6	573	12	US-10-425-114-67027	Sequence 67027, A
35	89.5	8.4	233	10	US-09-746-660A-52	Sequence 52, Appl
36	89.5	8.4	233	14	US-10-196-232-25	Sequence 25, Appl
37	89.5	8.4	236	9	US-09-738-626-6955	Sequence 6955, Ap
38	89.5	8.4	236	14	US-10-166-142-8	Sequence 8, Appli
39	87.5	8.2	465	12	US-10-282-122A-43620	Sequence 43620, A
40	87	8.2	364	12	US-10-424-599-247108	Sequence 247108,
41	85	8.0	484	9	US-09-815-242-10809	Sequence 10809, A
42	85	8.0	484	12	US-10-282-122A-56917	Sequence 56917, A
43	83.5	7.9	438	12	US-10-282-122A-61182	Sequence 61182, A
44	83.5	7.9	506	12	US-10-282-122A-49913	Sequence 49913, A
45	83.5	7.9	509	12	US-10-282-122A-49320	Sequence 49320, A

ALIGNMENTS

RESULT 1

US-09-927-395-2  
; Sequence 2, Application US/09927395  
; Patent No. US20020058314A1  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENIAMOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD  
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/927,395  
; CURRENT FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 09/396,357  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: RU98118425  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-927-395-2

Query Match 100.0%; Score 1061; DB 9; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.2e-110;  
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLEWFWFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCWAGRGV 60  
Db 1 MLEWFWFAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCWAGRGV 60  
QY 61 GTLFSRSVIAFEVLKAGAAAYLWIGIQOWRAACADLKLASTQSRRLHFORAVFVNLT 120

Db 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFVNL 120  
QY 121 NPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKGPKQ 180  
Db 121 NPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKGPKQ 180  
QY 181 MKALNKIFGSLFMLVGLLARSARHA 205  
Db 181 MKALNKIFGSLFMLVGLLARSARHA 205

RESULT 2  
US-09-847-392-2  
; Sequence 2, Application US/09847392  
; Patent No. US20020102670A1  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA  
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH  
; APPLICANT: BELAREOVA, ALL VALENTINOVNA  
; APPLICANT: TOKHAKOVA, IRINA LVOVNA  
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD  
; FILE REFERENCE: 0010-1039-0  
; CURRENT APPLICATION NUMBER: US/09/847,392  
; CURRENT FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 09/396,357  
; PRIOR FILING DATE: 1999-09-15  
; PRIOR APPLICATION NUMBER: RU98118425  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-847-392-2

Query Match 100.0%; Score 1061; DB 9; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.2e-110;  
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAGGVYCWASDRTGSDSYCAGWGV 60  
Db 1 MTEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAGGVYCWASDRTGSDSYCAGWGV 60  
QY 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFVNL 120  
Db 61 GTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFVNL 120  
QY 121 NPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKGPKQ 180  
Db 121 NPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKGPKQ 180  
QY 181 MKALNKIFGSLFMLVGLLARSARHA 205  
Db 181 MKALNKIFGSLFMLVGLLARSARHA 205

RESULT 3  
US-10-282-122A-60157  
; Sequence 60157, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282.122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 60157  
; LENGTH: 206  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-10-282-122A-60157

Query Match 78.3%; Score 830.5; DB 12; Length 206;  
Best Local Similarity 81.2%; Pred. No. 3.2e-84;  
Matches 169; Conservative 13; Mismatches 21; Indels 5; Gaps 3;  
QY 1 MTEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAGGVYCWASDRTG---DSYCAGW 57  
Db 1 MTEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAGGVYCWASDRTG---DSYCAGW 58  
QY 58 RGVGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV 117  
Db 59 VGLGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV 118  
QY 118 NLTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKG 177  
Db 119 NLTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTTIVVDIIIVMIGYATLAQRIALWIKG 178  
QY 178 PKQMKALNKIFGSLFMLVGLLARSARHA 205  
Db 179 PKQMKALNKIFGSLFMLVGLLARSARHA 206

RESULT 4  
US-10-282-122A-56211  
; Sequence 56211, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert



APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 56211  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Enterobacter cloacae  
US-10-282-122A-56211

Query Match 78.0%; Score 827.5; DB 12; Length 206;  
Best Local Similarity 79.1%; Pred. No. 7e-84;  
Matches 167; Conservative 15; Mismatches 18; Indels 11; Gaps 3;  
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRDGTGDSYCAGWR-- 58  
Db 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGY- RGAASIAAGLQTG---LGIHIV 55  
QY 59 ---GVGTLFPSRSVIAFEVLKWAGAAYLILWLGIOQWRAAGAILDKSLASTQSRRLFORA 114  
Db 56 LVGIGLGTLFPSRSVIAFEVLKWAGAAYLILWLGIOQWRAAGSINLTALTQNRHFLKRA 115  
QY 115 VFVNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTIIVDIIVMIGYATLAQRIALW 174  
Db 116 VFVNLTNPKSIVFLAALFPQFIVPHQPQVMQYVVLGATTIIVDIIVMIGYATLAQRIAAW 175  
QY 175 IKGPKQMKALNKIFGSLFMLVGALLASARHA 205  
Db 176 IKGPKQMKALNKVFGSLFMLVGALLASARHA 206

RESULT 5  
US-10-282-122A-75782  
Sequence 75782, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 75782  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Salmonella typhi  
US-10-282-122A-75782  
Query Match 77.1%; Score 818.5; DB 12; Length 206;  
Best Local Similarity 78.7%; Pred. No. 7e-83;  
Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;  
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRDGTGDSYCAGWR-- 58  
Db 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGY- RGAASIAAGLQTG---LGIHIV 55  
QY 59 ---GVGTLFPSRSVIAFEVLKWAGAAYLILWLGIOQWRAAGAILDKSLASTQSRRLFORA 114  
Db 56 LVGVGLGTLFPSRSVIAFEILKWAGAAYLILWLGIOQWRAAGAILDLHTLAQTQSRGLFKRA 115  
QY 115 VFVNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTIIVDIIVMIGYATLAQRIALW 174  
Db 116 IFVNLTNPKSIVFLAALFPQFIMPQPOLAQYLILGTTIIVDIIVMIGYATLAQRIAAW 175  
QY 175 IKGPKQMKALNKIFGSLFMLVGALLASARHA 205  
Db 176 IKGPKQMKALNKAFGSLFMLVGALLASARHA 206  
RESULT 6  
US-10-282-122A-72863  
Sequence 72863, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A

```

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed -
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72863
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (169)..(169)
; OTHER INFORMATION: X-any amino acid
US-10-282-122A-72863

```

[illegible]

RESULT 7  
US-10-282-122A-78024  
; Sequence 78024, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78024
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Yersinia pestis
; US-10-282-122A-78024

```

	Query Match	65.8%;	Score 698.5;	DB 12;	Length 206;
	Best Local Similarity	63.3%;	Pred. No. 1.7e-69;		
	Matches 136;	Conservative 33;	Mismatches 23;	Indels 23;	Gaps 3
Qy	1	MTLEWVFAVLTSIIILTSPGSGAINTMTTSLNHGVAGGVYCWASDRTGDSYCAGWR--	58		
Dd	1	MTLDWMLTYLLTTLILSLSPGSGAINTMSTAISHG--TRGVV-----ASIGGLQLG	49		
Qy	59	-----GVGTLPSSSVIAFEVLKWAGAAAYLIWLGIQQWRAAGAIDLKSLASTQSRR	108		
Dd	50	LAVHIVLVGVLGALVSQSLAFEILLKWLGAAAYLIWLGIQQWRAAGSLDLHALANSMPRR	109		
Qy	109	HLFQRADFVNLTNPKSIYFLAALFPQFTIMPQQPQLMQYIVLGVTTIVVDIIVMIGYATLA	168		
Dd	110	KLFKRADFVNLTNPKSIYFLAALFPQFVLPPQPQAQYLIGSTSVIVDIIVMIGYATLA	169		
Qy	169	QRIALWIKGPQMKNLKIFGSLFMLVGALLASAR	203		
Dd	170	TRIARWIKSQPMKLLNIFFGGLFMJGALLATAR	204		

```

RESULT 8
US-10-282-122A-42680
; Sequence 42680, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of E
; FILE REFERENCE: ELITRA.034A

```





QY 102 ASQSRRLHFRQAVFVNLTNPKSIVFLAALFPQFIMPPQPOLMOYIVLGVTTIVVDIIVM 161  
Db 105 RPVKPMTLVFRGFLVNISNPKALIFILAILPQFIEPTAPLEFMQYVIAATMVVDLIVM 164  
QY 162 IGYATLAQRIALWIKGPKOMKALNKIFGSLFMLVGLALAS 201  
Db 165 AGYGLASKVLKALKTPRQQRRLNRTFASLFFVGAAGFLAT 204

RESULT 13  
US-10-282-122A-67883  
; Sequence 67883, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Ttawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 67883  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: Pseudomonas putida  
US-10-282-122A-67883

Query Match 33.4%; Score 354.5; DB 12; Length 211;  
Best Local Similarity 32.3%; Pred. No. 4,1e-31;  
Matches 74; Conservative 41; Mismatches 69; Indels 45; Gaps 4;  
QY 1 MTEWFWAYLLTSILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWR-- 58  
Db 2 MSMEVNLGFFAACWVISLSLPGAGAIASMSSGLQYGF-----WRGY 41  
QY 59 -----GVGTLFSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAILDK 99  
Db 42 WNALGLQLGLIMOIAIAAGVGAVALAASATAFQVIKFWGVGLVYLAKQWRAL-PMDMS 100  
QY 100 SLASTQ---SRRLHFRQAVFVNLTNPKSIVFLAALFPQFIMPPQPOLMOYIVLGVTTIVV 156

Db 101 DESGVRRPIGKPLSLVFRGFLVNISNPKALVFMALVLPQFLNHPAPLLPOYVAITVTMTV 160  
QY 157 DIIVMTGYATLAQRIALWIKGPKOMKALNKIFGSLFMLVGLALASARHA 205  
Db 161 DLLVMAGYTGSLASHVLRMLRTPKQKRLNRTTAGLFIGAATELATLRR 209

RESULT 14  
US-10-282-122A-51102  
; Sequence 51102, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51102  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
US-10-282-122A-51102

Query Match 29.5%; Score 312.5; DB 12; Length 209;  
Best Local Similarity 33.5%; Pred. No. 2e-26;  
Matches 73; Conservative 39; Mismatches 79; Indels 27; Gaps 5;  
QY 1 MTEWFWAYLLTSILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGW--- 57  
Db 1 MTLSTWLTFFVVASWAIISFSPGAGAISAMSSGLKYGFARG---YWNT-----AGLILG 49  
QY 58 -----RGVGTLSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGA---IDLKSLASTQ 105  
Db 50 ILFQFLIVAVGLGAVLATSELAFTLVKYAGVVYLIYLGMRQIRTDAAPTVTDAGD-PHRA 108  
QY 106 SRRHLFRQAVFVNLTNPKSIVFLAALFPQFIMPPQPOLMOYIVLGVTTIVVDIIVMIGYA 165  
Db 109 SIRELVGRGFLINTMNPKGTVFLAVWPQFVDPAPQPLTQQYLALAGTLAFTDLVAMDVYT 168

QY 166 TLAQRIALWIKPKQKALNKIFGSLFMLVGLLASAR 203  
Db 169 LLAQVLRMLRKAHHRWNRVFGSLFILAGVFLATFR 206

Db 170 AALLSTVHQA 180

Search completed: March 24, 2004, 21:21:52  
Job time : 42 secs

RESULT 15

US-10-282-122A-44873  
; Sequence 44873, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 44873  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-10-282-122A-44873

Query Match 27.4%; Score 291; DB 12; Length 181;  
Best Local Similarity 33.5%; Pred. No. 4.1e-24;  
Matches 64; Conservative 37; Mismatches 66; Indels 24; Gaps 3;

QY 28 MTTSLNHGYPAGGVYCWASDRTGSDYCAGWR-----GVGTLFSRSVIAFEVLK 75  
Db 1 MSSGLNYGF-----RHGYWNAIGLQIALLIQIMIVAAGVGVLFATTPLAFAQVK 49  
QY 76 WAGAAYLIWLGIQWRA-AGAILDKSLASTQSRHFLQRAVFNLTNPKSTVFLAALFPQ 134  
Db 50 WFGVAYLLYLAYLQWTFAPVKDIEQHEKKDKSVSALLNGFVFNISNPKATVFLAVLPQ 109  
QY 135 FIMPQQQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKPKQKALNKIFGSLFML 194  
Db 110 FLDLSKPOWIOYLIIMATMTVIDLIWMAGYTGASKVRLRLFSRSPKQKYLNRGFAVMFSC 169  
QY 195 VGALLASARHA 205



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: March 24, 2004, 21:12:37 ; Search time 20 Seconds  
(without alignments)  
985.963 Million cell updates/sec

Title: US-09-847-392-2  
Perfect score: 1061  
Sequence: 1 MLEWFWFAYLLTSIILLTSP.....KIFGSLFMLVGLLASARHA 205  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	898.5	84.7	206	B91223	homoserine/homoser
2	818.5	77.1	206	AF0917	homoserine/homoser
3	698.5	65.8	206	AG0466	probable homoserin
4	686	64.7	138	A65187	hypothetical 15.4
5	680	64.1	138	H86069	hypothetical prote
6	459.5	43.3	205	G82358	conserved hypot het
7	389	36.7	209	C82990	hypothetical prote
8	256	24.1	212	AF2805	homoserine/homoser
9	256	24.1	273	F97584	hypothetical prote
10	245	23.1	210	AC2982	RhtB family transp
11	245	23.1	278	E98301	hypothetical prote
12	235	22.1	216	AH3203	RhtB family transp
13	228	21.5	241	C75329	conserved hypot het
14	208.5	19.7	203	AB2683	RhtB family transp
15	208.5	19.7	239	H97464	conserved hypot het
16	198	18.7	235	AD2904	RhtB family transp
17	198	18.7	235	F97679	hypothetical prote
18	195.5	18.4	212	F64940	hypothetical prote
19	195	18.4	210	F69975	dihydrodipicolinat
20	193	18.2	210	G83082	hypothetical prote
21	192.5	18.1	207	E83703	hypothetical prote
22	191.5	18.0	212	C90942	hypothetical prote
23	190	17.9	212	G82200	conserved hypot het
24	188	17.7	222	E87264	efflux protein, Ly
25	187.5	17.7	212	AD0714	probable membrane
26	186.5	17.6	212	G85790	hypothetical prote
27	186	17.5	208	C82471	conserved hypot het
28	182.5	17.2	214	AG2684	RhtB family transp
29	182.5	17.2	224	E97466	hypothetical prote

30	180.5	17.0	207	2	D83187	hypothetical prote
31	180.5	17.0	216	2	F83051	conserved hypot het
32	178.5	16.8	213	2	C82523	amino acid transpo
33	178.5	16.8	216	2	C55580	hypothetical prote
34	172	16.2	249	2	AG3485	homoserine/homoser
35	171	16.1	218	2	A96009	probable amino aci
36	168.5	15.9	210	2	E87252	efflux protein, Ly
37	167	15.7	212	2	H87498	lysE family transl
38	166.5	15.7	205	2	AH3641	transporter, lyse
39	166.5	15.7	208	2	G87305	efflux protein, Ly
40	166	15.6	206	2	C82139	conserved hypot het
41	163.5	15.4	205	2	AI2614	RhtB family transp
42	163.5	15.4	224	2	H97396	dihydrodipicolinat
43	161.5	15.2	210	2	AH2720	RhtB family transp
44	161.5	15.2	210	2	D97502	hypothetical prote
45	156	14.7	204	2	B83279	hypothetical prote

ALIGNMENTS

RESULT 1

B91223  
homoserine/homoserine lactone efflux protein [imported] - Escherichia coli (strain O15  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: B91223  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: B91223  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-206 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BAB38177.1; PID:gl3364230; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: ECs4754

Query Match 84.7%; Score 898.5; DB 2; Length 206;  
Best Local Similarity 89.9%; Pred. No. 7.3e-78;  
Matches 187; Conservative 3; Mismatches 13; Indels 5; Gaps 3;

QY	1	MTLEWFWFAYLLTSIILLTSPGSGAINTMTTSLNHGYPAGGVYCNASDRGTG---DSYCAGW	57
DB	1	MTLEWFWFAYLLTSIILLTSPGSGAINTMTTSLNHGY-RGAVASITAGLTGLAIHVLVG-	58
QY	58	RGVGTLSRSVIAFEVLKWAGAAYLWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV	117
DB	59	VGLGTLFSRSVIAFEVLKWAGAAYLWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFV	118
QY	118	NLTNPKSIVFLAALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKG	177
DB	119	NLTNPKSIVFLAALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKG	178
QY	178	PKQMKALNKIFGSLFMLVGLLASARHA	205
DB	179	PKQMKALNKIFGSLFMLVGLLASARHA	206

RESULT 2

AF0917  
homoserine/homoserine lactone efflux protein [imported] - Salmonella enterica subsp. en  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AF0917  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001



Db 1 MIAFEVLKWAGAAVLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFVNLTPKSI V 60

Qy 128 LAALFPQFIMPQPOLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKPKOMKALNKI 187

Db 61 LAALFPQFIMPQPOLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKPKOMKALNKI 120

Qy 188 FGSLEMLVGALLASARHA 205

Db 121 FGSLEMLVGALLASAXHA 138

RESULT 6

G82358

conserved hypothetical protein VC0136 [imported] - Vibrio cholerae (strain N16961 serog

C;Species: Vibrio cholerae

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: G82358

R;Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406933; PMID:10952301

A;Accession: G82358

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-205 <HEI>

A;Cross-references: GB:AE004104; GB:AE003852; NID:G9654534; PIDN:AAF93313.1; GSPDB:GN001

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC0136

A;Map position: 1

C;Superfamily: hypothetical protein b1798

Query Match 43.3%; Score 459.5; DB 2; Length 205;

Best Local Similarity 43.4%; Pred. No. 3.2e-36;

Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;

Qy 1 MLEWFWFAYLLTSIILTLSPGSGAINMTTSLNHG--YPAGGVYCWASDRGDSYCAGWR 58

Db 1 MDHVWLAAYLLTAVVFLAPGSGTNSISNGLSYGTRHSLGAILGLOGLACHIVLVG-I 59

Qy 59 GVGTLFSSRVIAFEVLKWAGAAVLIWLGIQQWRAAGAILDKSLASTQSRRLHFORAVFVN 118

Db 60 GIGALVAQSALAFLLIKWIGAAVLIWLGIQKWRDAPLTTATTSHELSQLLRKAVLIN 119

Qy 119 LTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTIVVDIIVMIGYATLAQRIALWIKGP 178

Db 120 LTNPKSIVFLVALFPQFIDPTRDHWQFLVLGITTITDAIVMFGYTALAAQLGRYIRSP 179

Qy 179 KQMKALNKIFGSLEMLVGALLASAR 203

Db 180 NIMTRMNLFGSMFMGCGMLLATAK 204

RESULT 7

C82990

hypothetical protein PA5249 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Feb-2002

C;Accession: C82990

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: C82990

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-209 <STO>

A;Cross-references: GB:AE004937; GB:AE004091; NID:G9951553; PIDN:AG08634.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA5249

C;Superfamily: hypothetical protein b1798

Query Match 36.7%; Score 389; DB 2; Length 209;

Best Local Similarity 36.4%; Pred. No. 1.6e-29;

Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;

Qy 1 MLEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58

Db 1 MLVSTWFAFFLACWALSISPGAGIASMSCGLQYGFARG-----YWNAL 44

Qy 59 -----GVGTLFSSRVIAFEVLKWAGAAVLIWLGIQQWRAAGAILDKSLAS 103

Db 45 GLQIGLALQIAIVAAGVALLATSALAFSLIKWFGVAYLVYLAVRQWQAP-----PQALS 99

Qy 104 TQSRRLHFQ-----RAVFNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLVGTTIVVD 157

Db 100 TGERPLGRPLTLVLGRFLVNASNPXAVIFMLAVLPQFIDPHQPLLAQYLIMGGTIVVD 159

Qy 158 IIVMIGYATLAQRIALWIKPKOMKALNKIFGSLEMLVGALLASARHA 205

Db 160 LIVWAGYTGLAARVLRVLRSPRQOKLVNRTFASLVFGAAGLLATVRR 207

RESULT 8

AF2805

homoserine/homoserine lactone efflux protein [imported] - Agrobacterium tumefaciens (st

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C;Accession: AF2805

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,

erage, G.; Gilliet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClel

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AF2805

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-212 <KUR>

A;Cross-references: GB:AE008688; PIDN:AAL42860.1; PID:G17740310; GSPDB:GN00186

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: rhtB

A;Map position: circular chromosome

C;Superfamily: hypothetical protein b1798

Query Match 24.1%; Score 256; DB 2; Length 212;

Best Local Similarity 34.1%; Pred. No. 6.7e-17;

Matches 71; Conservative 29; Mismatches 88; Indels 20; Gaps 6;

Qy 1 MLEWFWFAYLLTSIILTLSPGSGAINMTTSLNHGYPAG-----GVYCWASDRGDS--SY 53

Db 1 MPLENWLAFVAASAIMLAIPGPTILLVTSYALGHGRKASTATVTGV-----ALGDFETAM 54

Qy 54 CAGWRGVGTLFSSRVIAFEVLKWAGAAVLIWLGIQQWRA-----AGAILDKSLASTQSRRH 109

Db 55 TASMLGLGALLATSAAALFTGLKWIGAAVLIYLGIKLMRSPVGGEGA-DGQGVTRGRPLK 113

Qy 110 LFORAVFVNLTPKSIIVFLAALFPQFIMPQPOLMQYIVLVGTTIVVDIIVMIGYATLAQ 169

Db 114 IFLHAYIVTALNPKSIVFFVAFELPQFLVFTLPFPWPQVLFEATFLVLATVNAALYGLLAS 173

Qy 170 RIALWIKPKOMKALNKIFGSLEMLVGA 197

Db 174 AARTIRKPKVQIRIVNRTGGGL--LIGA 199

RESULT 9



2Y 9 YLLTSLIILTLSPGSGAINTMTTSLNHGYAG-----GVYCWASDRDTGDSYACAGWRGVGTL 63  
Db 14 FISTTLVLLVPGPAVLXIFARSVEQGRSAGLSILGIH---TATLVHVAAAVGLSAL 69  
2Y 64 FRSRVIAFEVLKWAGAAAYLIWLGIQOWRAAGAI-DLKSLASTQSRRLHFORAVFVNLTNP 122  
Db 70 LASSALAFSVVKYAGAAAYLIWLGLKLFGPSDIPDVEGGLPTRSRMRIFREGFIVNLLNP 129  
2Y 123 KSIIVFLAALFPQFIMPQPOL-MQYIVLGVTTIVVDIIVMIGYATLAQRIALWIK-GPKQ 180  
Db 130 KTALFFLAFLPQFVEVDRGHVAMQIAFLGILYTAIGILTDTGYALVAGTAGNWLKRSPVY 189  
2Y 181 MKALNKIFGSLFMLVGALLASA 202  
Db 190 LKAERWVSGFVYIGLVTAFA 211  
RESULT 13  
C75329  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Jun-2002  
C;Accession: C75329  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: C75329  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-241 <WHI>  
A;Cross-references: GB:AE002038; GB:AE000513; NID:G6459790; PIDN:AAF11548.1; PID:G645978  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR1999  
A;Map position: 1  
C;Superfamily: hypothetical protein b1798  
Query Match 21.5%; Score 228; DB 2; Length 241;  
Best Local Similarity 30.3%; Pred. No. 3.5e-14;  
Matches 64; Conservative 41; Mismatches 84; Indels 22; Gaps 6;  
2Y 8 AYLLTSIILTLSPGSGAINTMTTSLNHGYAGGVYCWASDRDTGDSYACAGWR-----G 59  
Db 38 AFLVAAVVLLALPGLMYILARSLGGR-----WAGIQSALGTGAGMVHVLASAVG 90  
2Y 60 VGTLFPSRVIAFEVLKWAGAAAYLIWLGIQOWRAAGAI-DLK---SLASTQSRRLHFORA 114  
Db 91 LSALIMASSLAFSVVKYAGAAAYLIYGLRVLLSKEALSSKEEASLAAAAPERQSRFLTQ 150  
2Y 115 VFVNLTNPKSIVFLAALFPQFIMPQPOLM-QYIVLGVTTIVVDIIVMIGYATLAQRIAL 173  
Db 151 AMTELLNPKTALFFLAVIPQFVKPATGHVFGQFLLGLTTSVVVNTLNAMLVATLAGFLGA 210  
2Y 174 WIKG-PKQMKALNKIFGSLFMLVGALLASAR 203  
Db 211 RLQGNPRFORQKQVASGGAMIALGTYYAVER 241  
RESULT 14  
AB2683  
rhtB family transporter rhtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AB2683  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
i Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AB2683  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-203 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL41880.1; PID:g17739243; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: rhtB  
A;Map position: circular chromosome  
C;Superfamily: hypothetical protein b1798  
Query Match 19.7%; Score 208.5; DB 2; Length 203;  
Best Local Similarity 26.6%; Pred. No. 2e-12;  
Matches 54; Conservative 37; Mismatches 93; Indels 19; Gaps 4;  
QY 9 YLLTSIILTLSPGSGAINTMTTSLNHGYAGGVYCWASDRDTGDSYC-----AGWRG 59  
Db 5 FLITSFIVAASPGTGVVYTLAAGLSQGAASIIAAG-----CTLGIVPHLLAAITG 56  
QY 60 VGTLFPSRVIAFEVLKWAGAAAYLIWLGIQOWRAAGAI-DLKSLASTQSRRLHFORAVFVNL 119  
Db 57 LAAILHTSALAFGIVKYLGVAYLLYMAWNTLRENGALKIDETROPQKPARVIGEAILINL 116  
QY 120 TNPKSIVFLAALFPQFIMP-QQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW-IGK 177  
Db 117 LNPKLSIFFFAFLPQFIAPGELSPTWRMIDLGLIFMALTFVLVYALYGCFGAASVRRQVVSR 176  
QY 178 PKQMKALNKIFGSLFMLVGALLA 200  
Db 177 PAVLAWLRRSFAAAFVALGAKLA 199  
RESULT 15  
H97464  
conserved hypothetical protein BH0429 (AP001508) [imported] - Agrobacterium tumefaciens  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: H97464  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, P.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: H97464  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-239 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK86673.1; PID:g15155857; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_1583  
A;Map position: circular chromosome  
Query Match 19.7%; Score 208.5; DB 2; Length 239;  
Best Local Similarity 26.6%; Pred. No. 2.4e-12;  
Matches 54; Conservative 37; Mismatches 93; Indels 19; Gaps 4;  
QY 9 YLLTSIILTLSPGSGAINTMTTSLNHGYAGGVYCWASDRDTGDSYC-----AGWRG 59  
Db 41 FLITSFIVAASPGTGVVYTLAAGLSQGAASIIAAG-----CTLGIVPHLLAAITG 92  
QY 60 VGTLFPSRVIAFEVLKWAGAAAYLIWLGIQOWRAAGAI-DLKSLASTQSRRLHFORAVFVNL 119  
Db 93 LAAILHTSALAFGIVKYLGVAYLLYMAWNTLRENGALKIDETROPQKPARVIGEAILINL 152  
QY 120 TNPKSIVFLAALFPQFIMP-QQPQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW-IGK 177  
Db 153 LNPKLSIFFFAFLPQFIAPGELSPTWRMIDLGLIFMALTFVLVYALYGCFGAASVRRQVVSR 212  
QY 178 PKQMKALNKIFGSLFMLVGALLA 200

Db 213 PAVLAWLRRSFVAAAFVALGAKLA 235

Search completed: March 24, 2004, 21:16:42  
Job time : 21 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:08:46 ; Search time 18 Seconds  
(without alignments)  
593.021 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MTLEWFWFAYLTSIILTSP.....KIFGSLFMLVGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	898.5	84.7	206	1 RHTB_ECOLI	P27847 escherichia
2	819.5	77.2	206	1 RHTB_SALTY	Q916n6 salmonella
3	818.5	77.1	206	1 RHTB_SALTY	Q823b4 salmonella
4	195.5	18.4	212	1 YEAS_ECOLI	P75249 escherichia
5	195	18.4	210	1 YRHP_BACSU	O05406 bacillus su
6	180.5	17.0	216	1 YBF7_PSEAE	P38102 pseudomonas
7	147	13.9	206	1 YG27_SYNY3	P74343 synechocyst
8	134	12.6	206	1 RHTC_SALTY	Q823b3 salmonella
9	134	12.6	206	1 RHTC_SALTY	Q916n7 salmonella
10	126.5	11.9	223	1 YAHN_ECOLI	P75693 escherichia
11	126	11.9	195	1 YFIK_ECOLI	P38101 escherichia
12	122	11.5	206	1 RHTC_ECOLI	P27846 escherichia
13	116	10.9	203	1 CHPE_PSEAE	O87005 pseudomonas
14	113	10.7	211	1 YGGA_ECOLI	P11667 escherichia
15	111.5	10.5	206	1 YGGA_AERSA	P70775 aeromonas s
16	94	8.9	457	1 PROY_ECOLI	P77327 escherichia
17	92	8.7	228	1 LYSE_COREF	Q8rqm4 corynebacte
18	92	8.7	456	1 PROY_SALTY	P37460 salmonella
19	89.5	8.4	233	1 LYSE_CORGL	P94633 corynebacte
20	87	8.2	210	1 YD07_HAEIN	Q57320 haemophilus
21	81	7.6	732	1 ATZN_ECOLI	P37617 escherichia
22	79.5	7.5	384	1 CYNX_ECOLI	P17583 escherichia
23	79.5	7.5	581	1 FUR4_SCHPO	Q10279 schizosacch
24	79	7.4	415	1 NQRB_VIBCH	Q9kps2 vibrio chol
25	79	7.4	615	1 NTDO_CAEEL	Q03614 caenorhabdi
26	78.5	7.4	366	1 SPSE_BACSU	P07373 bacillus su
27	78.5	7.4	471	1 PPOX_MYXXA	P56601 myxococcus
28	78	7.4	1694	1 SN_MOUSE	Q62230 mus musculu
29	77.5	7.3	225	1 YGGA_AERHY	P52047 aeromonas h
30	77	7.3	348	1 OPSD_SARXA	P79914 sargocentro
31	77	7.3	543	1 SGLT_VIBPA	P96169 vibrio para
32	76	7.2	195	1 Y688_RHILO	Q98m84 rhizobium l
33	76	7.2	284	1 MODD_ECO57	P58496 escherichia

ALIGNMENTS

RESULT 1

ID	RHTB_ECOLI	STANDARD;	PRT;	206 AA.
AC	P27847;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Homoserine/homoserine lactone efflux protein.			
GN	RHTB OR B3824 OR C4746 OR Z5345 OR ECS4754 OR SF3902 OR S3853.			
OS	Escherichia coli, O6,			Q8ka37 buchnera ap
OS	Escherichia coli O157:H7, and			Q01625 bacillus su
OS	Shigella flexneri.			Q64252 homo sapien
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			P77429 escherichia
OC	Enterobacteriaceae; Escherichia.			P53387 kluyveromyc
OX	NCBI_TaxID=562, 217992, 83334, 623;			P05715 thermoprote
RN	[1]			Q10858 mycobacteri
RP	SEQUENCE FROM N.A.			P45064 haemophilus
RC	SPECIES=E.coli; STRAIN=K12 / MG1655;			P11631 oncorhynchu
RX	MEDLINE=92358234; PubMed=1379743;			Q91049 gallus gall
RA	Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;			P28873 candida alb
RT	"Analysis of the Escherichia coli genome: DNA sequence of the region			O15040 homo sapien
RT	from 84.5 to 86.5 minutes."			
RL	Science 257:771-778(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;			
RX	MEDLINE=22388234; PubMed=12471157;			
RA	Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,			
RA	Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,			
RA	Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,			
RA	Mobley H.L.T., Donnenberg M.S., Blattner F.R.;			
RT	"Extensive mosaic structure revealed by the complete genome sequence			
RT	of uropathogenic Escherichia coli."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoussis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."			
RL	Nature 409:529-533(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;			
RX	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli			



Query Match 77.2%; Score 819.5; DB 1; Length 206;  
Best Local Similarity 79.2%; Pred. No. 3.2e-69;  
Matches 165; Conservative 15; Mismatches 20; Indels 11; Gaps 3;  
  
QY 1 MTLFWFAYLLTSILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRDTSYCAGWR-- 58  
DB 1 MTFEWFAYLLTSILTLSPGSGAINTMTTSLNHGY-RGAVASIAGLQTG---LGIHIV 55  
  
QY 59 ---GVGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDKSLASTQSRRLHFORA 114  
DB 56 LVGVGLGTLFSSRSVIAFEILKWAGAAAYLIWLGIQWRAAGAIIDHTLAQTQSRGLFKRA 115  
  
QY 115 VFVNLTPKSIIVFLAALFPQFIMPQQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 174  
DB 116 IFVNLTPKSIIVFLAALFPQFIMPQQLAQYLILGVTTIVVDVMTGYATLAQRIAAW 175  
  
QY 175 IKGPQMKALNKIFGSLFMLVGALLASARHA 205  
DB 176 IKGPQMKALNKAFGSLFMLVGALLASARHA 206

RESULT 3  
RHTB\_SALTI  
ID\_RHTB\_SALTI STANDARD; PRT; 206 AA.  
AC Q823B4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homoserine/homoserine lactone efflux protein.  
GN RHTB OR STY3599 OR T3337.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -!- FUNCTION: Conducts the efflux of homoserine and homoserine  
lactone (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the rht family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL; AL627278; CAD07932.1; -.

DR EMBL; AB016845; AAO70865.1; -.  
DR InterPro; IPR004778; Homoser\_Thr\_eff.  
DR InterPro; IPR001123; Lyse.  
DR Pfam; PF01810; Lyse; 1.  
DR TIGRFAMs; TIGR00949; 2A76; 1.  
KW Transport; Transmembrane; Complete proteome.  
FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 45 65 POTENTIAL.  
FT TRANSMEM 68 88 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
SQ SEQUENCE 206 AA; 22308 MW; 40940D3DC0A5CAF2 CRC64;  
  
Query Match 77.1%; Score 818.5; DB 1; Length 206;  
Best Local Similarity 78.7%; Pred. No. 4e-69;  
Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;  
  
QY 1 MTLFWFAYLLTSILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRDTSYCAGWR-- 58  
DB 1 MTFEWFAYLLTSILTLSPGSGAINTMTTSLNHGY-RGAAASIAGLQTG---LGIHIV 55  
  
QY 59 ---GVGTLFSSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDKSLASTQSRRLHFORA 114  
DB 56 LVGVGLGTLFSSRSVIAFEILKWAGAAAYLIWLGIQWRAAGAIIDHTLAQTQSRGLFKRA 115  
  
QY 115 VFVNLTPKSIIVFLAALFPQFIMPQQLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 174  
DB 116 IFVNLTPKSIIVFLAALFPQFIMPQQLAQYLILGVTTIVVDVMTGYATLAQRIAAW 175  
  
QY 175 IKGPQMKALNKIFGSLFMLVGALLASARHA 205  
DB 176 IKGPQMKALNKAFGSLFMLVGALLASARHA 206

RESULT 4  
YEAS\_ECOLI  
ID\_YEAS\_ECOLI STANDARD; PRT; 212 AA.  
AC P76249; O07969; O07971;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein yeas.  
GN YEAS OR B1798.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
RA Yamamoto Y., Horiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map.";  
RL DNA Res. 3:379-392(1996).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the rht family.







QY 71 --FEVLKWAGAAVLIWLGIQWRA-----AGAIDLKSLASTQSRRLHFQRAVFNLTNP 122  
Db 73 DFFVILKYISGIYLSWLGINTIRAKVNNQSLAKVDKSLSS-----FSAGLLITLADQ 126  
QY 123 KSIVFLAALFPQFIMPQQLMQYIVLGVTTIVVDIIIVMIGVATLAQRIALWIKPKQMK 182  
Db 127 KAVLPYLGFLPTFEVDVNNIAYLDIAVILITALTGVGVKIFYAFLAHRSGLLI-SRQNK 185  
QY 183 ALNKIFGSLFMLVGALL 199  
Db 186 IMNYLAGALMISGVFL 202

RESULT 8

RHTC\_SALTI  
ID RHTC\_SALTI STANDARD; PRT; 206 AA.  
AC Q8Z3B3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Threonine efflux protein.  
GN RHTC OR STY3600 OR T3338.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
RT and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the rht family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL627278; CAD07933.1; -.  
CC EMBL; AE016845; AA070866.1; -.  
CC InterPro; IPR004778; Homoser\_Thr\_eff.  
CC InterPro; IPR001123; Lyse.  
CC Pfam; PF01810; Lyse; 1.  
CC TIGRFAMs; TIGR00949; 2A76; 1.  
CC Transport; Transmembrane; Complete proteome.  
KW TRANSMEM 1 21 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 67 87 POTENTIAL.  
FT TRANSMEM 150 173 POTENTIAL.

SQ SEQUENCE 206 AA; 22480 MW; E9DFCDF4E42D46A0 CRC64;  
Query Match 12.6%; Score 134; DB 1; Length 206;  
Best Local Similarity 22.1%; Pred. No. 1.9e-05;  
Matches 50; Conservative 32; Mismatches 92; Indels 52; Gaps 8;  
QY 6 WFAVLLTSIIILTLSPGSGAINTMTTSLNHGYPAG-----GVYCWASDRTGDSYCAGWRGV 60  
Db 5 FFTVAMVHIVALMSPGPDFFVSVQTAVSRSRKEAMMGVLGTCGV-----MVWAGV 55  
QY 61 GTLFSRSVIAFEVLKW-----AGAAVLIWLGIQWRAAGAILDKSLASTQSRRLH--- 110  
Db 56 ALLGLHLII--EKVAVLHTIIMVGGGLYLCWVGQMLRGA----LKKQDAASSPHIELA 109  
QY 111 -----FQRAVFNLTNPKSIIVFLAALFPQFIMPQQLMQYIVLGVTTIVVD 157  
Db 110 QSGRSFLKGLLTNLSNPKAIIYFGSVFLFVGDVNVGAAARWGIFALITLETIAWFTVVAS 169  
QY 158 IIVMIGVATLAQRIALWIKPKQKALNKIFGSLFMLVGLLALASAR 203  
Db 170 LFALPKVRRGYQRLAKWIDG-----FAGALFAGFGIHLIISR 206

RESULT 9

RHTC\_SALTY  
ID RHTC\_SALTY STANDARD; PRT; 206 AA.  
AC Q9L6N7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Threonine efflux protein.  
GN RHTC OR STM3959 OR STMD1.31.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the rht family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF233324; AAF33433.1; -.  
CC EMBL; AE008884; AAL22803.1; -.  
CC StyGene; SG????; rhtC.  
CC InterPro; IPR004778; Homoser\_Thr\_eff.  
CC InterPro; IPR001123; Lyse.  
CC Pfam; PF01810; Lyse; 1.  
CC TIGRFAMs; TIGR00949; 2A76; 1.  
CC Transport; Transmembrane; Complete proteome.  
KW TRANSMEM 1 21 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 57 87 POTENTIAL.  
FT TRANSMEM 150 173 POTENTIAL.  
SQ SEQUENCE 206 AA; 22498 MW; 080477853FC2733F CRC64;







```
DR EMBL; M30198; -; NOT_ANNOTATED_CDS.
DR PIR; A98223; A98223.
DR EcoGene; EG11468; rhtC.
DR InterPro; IPR004778; Homoser_Thr_eff.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
DR TIGRFAMs; TIGR00949; 2A76; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22474 MW; F64017878CC6D50D CRC64;

Query Match 11.5%; Score 122; DB 1; Length 206;
Best Local Similarity 21.0%; Pred. No. 0.00025;
Matches 46; Conservative 36; Mismatches 89; Indels 48; Gaps 8;

QY 11 LTSIIILTLSPGSGAINTMTTSLNHGYPAG-----GVYCWASDRTGDSYCACWGRGVGTLPFS 65
Db 10 MVHIVALMSPGDPFFVSQTAVSRSRKEAMGMVLGITCGV-----MVWAGIALLLGL 60

QY 66 RSVIAFEVLKW-----AGAAYLWLGIQWRAAGAILDKSLASTQSRRLH-----FQ 112
Db 61 HLII--EKMAWLHTLIMVGGGLYLCWMGYQMLR--GALKKEAVSAPAPQVELAKSGRSFL 116

QY 113 RAVFVNLTNPKSIVFLAALFPQFIMPQQLMQVIVLGV-----TIVVDIIVMIGY 164
Db 117 KGLLTNLNPKAIIIVFGSVFLFVGDNVGTTARWGIFALIIIVETLAWFTVVASLFALPQM 176

QY 165 ATLAQRIALWIKGPKQMKALNKIFGSLFMLVGLLASAR 203
Db 177 RRGYQRLAKWIDG-----FAGALFAGFGIHLIISR 206
```

```
RESULT 13
CHPE_PSEAE
ID_CHPE_PSEAE STANDARD; PRT; 203 AA.
AC O87005;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chemotactic transduction protein chpE.
GN CHPE OR PA0417.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Whitcomb C.B., Young M.D., Hobbs M., Mattick J.S.;
RT "Pseudomonas aeruginosa chemotactic transduction genes pill, chpA
RT chpB and downstream genes chpC, chpD and chpE."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the rht family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U79580; AAC23935.1; -.
DR EMBL; AE004479; AAG03806.1; -.
DR PIR; T30317; T30317.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
KW Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
SQ SEQUENCE 203 AA; 21290 MW; 195553C048AAD099 CRC64;

Query Match 10.9%; Score 116; DB 1; Length 203;
Best Local Similarity 30.2%; Pred. No. 0.0009;
Matches 39; Conservative 21; Mismatches 63; Indels 6; Gaps 4;

QY 6 WFAYLLTSLTLSPGSGAINTMTTSLNHGY-PAGGVYCWASDRTGDSYCA--GWRGVGT 62
Db 5 FLAALLFGFAPNVSPGAVFSETLRRGLTGGRPA--LLVQLSLIGDAVWALLGLTGLAL 62

QY 63 LFSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAILDKSLASTQSRRLHFORAVFVNLTNP 122
Db 63 LLGYEQVRIP-LTLCAAYLAWLGVQGLRDWSPPLAEDAGEQGRNFAFGAGAAISLNP 121

QY 123 KSIVFLAAL 131
Db 122 KNVVYWGAL 130
```

```
RESULT 14
YGGG_ECOLI
ID_YGGG_ECOLI STANDARD; PRT; 211 AA.
AC P11667;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yggA.
GN YGGG OR B2923.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
SEQUENCE OF 15-211 FROM N.A.
RC STRAIN=K12 / CS520;
RX MEDLINE=89313302; PubMed=2546007;
RA Alefounder P.R., Baldwin S.A., Perham S.A., Short N.J.;
RT "Identification, molecular cloning and sequence analysis of a gene
RT cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-
RT phosphate dehydrogenase and a putative second glyceraldehyde 3-
RT phosphate dehydrogenase of Escherichia coli."
RL Mol. Microbiol. 3:723-732(1989).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: March 24, 2004, 21:12:07 ; Search time 45 Seconds  
(without alignments)  
1437.361 Million cell updates/sec

Title: US-09-847-392-2  
Perfect score: 1061  
Sequence: 1 MLEWFWFAYLLTSIILTSP.....KIFGSLFMLVGALLASARHA 205

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	698.5	65.8	206	16	Q8ZAH0 yersinia pe
2	470.5	44.3	207	16	Q8DD72
3	459.5	43.3	205	16	Q9KVK7 vibrio chol
4	457.5	43.1	205	16	Q87KK4 vibrio para
5	389	36.7	209	16	Q9HTU5 pseudomonas
6	367.5	34.6	210	16	Q88B79 pseudomonas
7	354.5	33.4	210	16	Q88RD4 pseudomonas
8	312.5	29.5	209	16	Q7W0G4 bordetella
9	311.5	29.4	209	16	Q7WDS2 bordetella
10	311.5	29.4	209	16	Q7W2S4 bordetella
11	294	27.7	203	16	Q88H00 pseudomonas
12	286.5	27.0	219	16	Q87PX1 vibrio para
13	281	26.5	211	16	Q92NK8 rhizobium m
14	269	25.4	207	16	Q8EKH1 shewanella
15	258	24.3	208	2	Q8KW80 ruegeria sp
16	256	24.1	273	16	Q8UE96 agrobacteri

17	255.5	24.1	204	16	Q98GT1 rhizobium l
18	252.5	23.8	209	16	Q7WQR5 bordetella
19	252.5	23.8	209	16	Q7WIT8 bordetella
20	248	23.4	202	16	Q7W4X4 bordetella
21	248	23.4	207	16	Q8XCZ0 chlorobium m
22	247	23.3	211	16	Q92LY9 rhizobium m
23	245	23.1	278	16	Q8UAB4 agrobacteri
24	244.5	23.0	209	16	Q7VS75 bordetella
25	241.5	22.8	218	16	Q7WGF2 bordetella
26	235.5	22.2	208	16	Q983T1 rhizobium l
27	235	22.1	209	16	Q887R5 pseudomonas
28	235	22.1	234	16	Q8UJW7 agrobacteri
29	228.5	21.5	207	16	Q98CS3 rhizobium l
30	228	21.5	241	16	Q9RSX2 deinococcus
31	222	20.9	213	16	Q88GU2 pseudomonas
32	221.5	20.9	208	16	Q89J33 bradyrhizob
33	218.5	20.6	203	16	Q98MB1 rhizobium l
34	214.5	20.2	203	16	Q7WJY4 bordetella
35	214.5	20.2	203	16	Q7WAS5 bordetella
36	212.5	20.0	203	16	Q7VVP8 bordetella
37	212	20.0	220	16	Q8XZ84 ralstonia s
38	211.5	19.9	206	16	Q88IH4 pseudomonas
39	210.5	19.8	211	17	Q8PSF6 methanosarc
40	210	19.8	210	16	Q81QR4 bacillus an
41	209.5	19.7	206	16	Q8Y2B8 ralstonia s
42	208.5	19.7	239	16	Q8UH19 agrobacteri
43	208	19.6	210	16	Q81DR4 bacillus ce
44	206	19.4	208	16	Q88KW8 pseudomonas
45	205.5	19.4	203	16	Q89DU7 bradyrhizob

ALIGNMENTS

RESULT 1  
Q8ZAH0  
ID Q8ZAH0 PRELIMINARY; PRT; 206 AA.  
AC Q8ZAH0;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Putative homoserine/homoserine lactone efflux protein.  
GN RHTB OR YPO3831 OR Y0399.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIM5 / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
DR EMBL; AJ414159; CAC93299.1; -.  
DR EMBL; AB013640; AAM83988.1; -.







Db 45 GLQIALVAQIAVAAGLGAVALAASEMAFTLIKWFGVAYLVILGKQWRASPTDLADESAV 104

QY 102 ASTQSRRLHFORAVFVNLTNPKSIVFLAALFPQFIMPQQPQLMQYIVLGVTTIVVDIIVM 161

Db 105 RPVGKPMTLVFRGFLVNIISNPKALIFILAILPQFIEPTAPLFMQYVILAAATWVVVDLIVM 164

QY 162 IGYATLAQRALWIKGPKQMKALNKIFGSLFMLVGALLAS 201

Db 165 AGYGLASKVLRLAKTPQQRRLNRTFASLVFGAAGFLAT 204

RESULT 7

Q88RD4 PRELIMINARY; PRT; 210 AA.

AC Q88RD4;

DT 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Transporter, Lyse family.

GN PP0198.

OS Pseudomonas putida (strain KT2440).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=160488;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22423060; PubMed=12534463;

RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,

RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,

RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,

RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,

RA Chris Lee P., Holtzappple E., Scanlan D., Tran K., Moazzez A.,

RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,

RA Lauber J., Stjepandic D., Hohnel J., Straetz M., Heim S.,

RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,

RA Fraser C.M.;

RT "Complete genome sequence and comparative analysis of the

RT metabolically versatile Pseudomonas putida KT2440.";

RL Environ. Microbiol. 4:799-808(2002).

DR EMBL; AE016774; AAN65831.1; -.

DR TIGR; PP0198; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005293; F:lysine permease activity; IEA.

DR GO; GO:0006865; P:amino acid transport; IEA.

DR InterPro; IPR001123; Lyse.

DR Pfam; PF01810; Lyse; 1.

KW Complete proteome.

SQ SEQUENCE 210 AA; 22827 MW; 89E800A87428770A CRC64;

Query Match 33.4%; Score 354.5; DB 16; Length 210;

Best Local Similarity 32.3%; Pred. No. 2.2e-25;

Matches 74; Conservative 41; Mismatches 69; Indels 45; Gaps 4;

QY 1 MTLWWFAYLLTSIIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGWR-- 58

Db 1 MSMEVWLGFFAACWVISLSPGAGAIASMSGLOQYGF-----WRGY 40

QY 59 -----GVGTLFSRSVIAFEVLKWAGAAVLIWLGIQWRAGAIIDLK 99

Db 41 WNALGLQLIMQIAIIAAGVGAVLAASATAFQIKWFGVGLVYLAYKQWRAL-PYDMS 99

QY 100 SLASTQ---SRRHLFORAVFVNLTNPKSIVFLAALFPQFIMPQQPQLMQYIVLGVTTIVV 156

Db 100 DESGVRPIGKPLSLVFRGFLVNIISNPKALVFMLAVLPQFLNPHAPLLPQYVAITVTMTIV 159

QY 157 DIIVMIGYATLAQRALWIKGPKQMKALNKIFGSLFMLVGALLASARHA 205

Db 160 DLLVMAGYTGASHVLEMLRTPKQQRRLNRTFAGLFIGAATFLATLRRRA 208

RESULT 8

Q7WOG4 PRELIMINARY; PRT; 209 AA.

ID Q7WOG4

QY 01-OCT-2003 (TReMBLrel. 25, Created)

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)

DE Lyse family efflux protein.

GN BB4915.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.

OX NCBI\_TaxID=518;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RB50 / ATCC BAA-588;

RX MEDLINE=22827954; PubMed=12910271;

RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,

RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,

AC Q7WOG4;

DT 01-OCT-2003 (TReMBLrel. 25, Created)

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Lyse family efflux protein.

GN BP0171.

OS Bordetella pertussis.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.

OX NCBI\_TaxID=520;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;

RX MEDLINE=22827954; PubMed=12910271;

RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,

RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,

RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,

RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,

RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,

RA Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;

RT "Comparative analysis of the genome sequences of Bordetella pertussis,

RT Bordetella parapertussis and Bordetella bronchiseptica.";

RL Nat. Genet. 35:32-40(2003).

DR EMBL; BX640411; CAB40550.1; -.

KW Complete proteome.

SQ SEQUENCE 209 AA; 22779 MW; F77DDE3CDD41E54 CRC64;

Query Match 29.5%; Score 312.5; DB 16; Length 209;

Best Local Similarity 33.5%; Pred. No. 2e-21;

Matches 73; Conservative 39; Mismatches 79; Indels 27; Gaps 5;

QY 1 MTLWWFAYLLTSIIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAGW--- 57

Db 1 MTLSTWLTFFVASWAIISPSGAGAISAMSSGLKYGFARG---YWNVT-----AGLILG 49

QY 58 -----RGVTLFSRSVIAFEVLKWAGAAVLIWLGIQWRAGAA---IDLKSLASTQ 105

Db 50 ILFQFLIVAVGLGAVLATSELAFTLVKIAGVVLYLYLGMRIRTDAAPVTVDAGD-PHRA 108

QY 106 SRRHLFORAVFVNLTNPKSIVFLAALFPQFIMPQQPQLMQYIVLGVTTIVVDIIVMIGYA 165

Db 109 SIRELVGRGFLINTMNPKGTVFLLAVVPQFVDPAPQPLTQQYLALAGTLAFTDLVAMDVYT 168

QY 166 TLAQRALWIKGPKQMKALNKIFGSLFMLVGALLASAR 203

Db 169 LLAAKVLRMLRKAHHRWNRVFGSLFILAGVFLATFR 206

RESULT 9

Q7WDS2 PRELIMINARY; PRT; 209 AA.

AC Q7WDS2;

DT 01-OCT-2003 (TReMBLrel. 25, Created)

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)

DE Lyse family efflux protein.

GN BB4915.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Alcaligenaceae; Bordetella.

OX NCBI\_TaxID=518;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RB50 / ATCC BAA-588;

RX MEDLINE=22827954; PubMed=12910271;

RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,

RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,



```

[1]
SEQUENCE FROM N.A.
RN RP
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sincorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591789; CAC46770.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
KW Complete proteome.
SQ SEQUENCE 211 AA; 22293 MW; 7955F42044DBCF47 CRC64;

Query Match 26.5%; Score 281; DB 16; Length 211;
Best Local Similarity 33.8%; Pred. No. 1.8e-18;
Matches 71; Conservative 27; Mismatches 106; Indels 6; Gaps 3;

QY 1 MTEWFWFAYLLTSIIITLSPGSGAINMTTSLNHGYPAGGVYCWASDRDGTG--SYCAGWR 58
Db 1 MSFEHWFAPAAASAVLLAIPGPTILLVISYALGHGRKIAGATV-AGVALGDFTAATASML 59
QY 59 GVGTLFRSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDLKSLASTQSR--HLFQRAV 115
Db 60 GLGALLATSAAVFTVLKWIAGAAAYLVWLGIKLRAPVCGNDSGTSTVETSPAERPLRIFLHTY 119
QY 116 FVNLTPKSIIVFLAALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYATLAORIALWI 175
Db 120 AVTALNPKSILFFVAFDLPQFLDLSRPLPAQMAIFETTFLLILATINAALYAWLAAAAGSTI 179
QY 176 KGPQKMKALKNKIFGSLFMLVGALLASARHA 205
Db 180 RKPNIIRIVNRLGGSLIGAGFLTAGLKRA 209

RESULT 14
Q8EKH1 PRELIMINARY; PRT; 207 AA.
ID Q8EKH1
AC Q8EKH1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DN Transporter, Lyse family.
GN S00122.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015463; AAN53209.1; -
DR TIGR; S00122; -
DR GO; GO:0016020; C:membrane; IEA.

RESULT 12
Q87PX1 PRELIMINARY; PRT; 219 AA.
AC Q87PX1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative homoserine/homoserine lactone efflux protein.
GN VP1379.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
DR EMBL; AP005077; BAC59642.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF01810; Lyse; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
KW Complete proteome.
SQ SEQUENCE 219 AA; 23558 MW; FBED455165D57500 CRC64;

Query Match 27.0%; Score 286.5; DB 16; Length 219;
Best Local Similarity 29.7%; Pred. No. 5.8e-19;
Matches 65; Conservative 43; Mismatches 86; Indels 25; Gaps 3;

QY 1 MTEWFWFAYLLTSIIITLSPGSGAINMTTSLNHGYPAGGVYCWASDRDGTGDSYCAG---- 56
Db 12 MQLDTWIYTLAILVLTASPGPSSLLCLSKGVSSGF-----RLALTALGSLSA 60
QY 57 -----WRGVGTLFRSRSVIAFEVLKWAGAAAYLIWLGIQWRAAGAIIDLKSLASTQSR 108
Db 61 ITIILTSFTGLGVVTSSEFVFNIIKWCAGAAAYLIWLGIQAFRSKQNDFAKSDSAQVSTS 120
QY 109 HL--FQRAVFNLTNPKSIVFLAALFPQFIMPQQLMQYIVLVGTTIVVDIIVMIGYAT 166
Db 121 HVSAYTSGFIVGSSNPKAIIFFTALEPQFIDPTASLLTQYAIFAGTFVVFELSWLTFYAL 180
QY 167 LAQRIALWIKGPKQMKALKNKIFGSLFMLVGALLASARHA 205
Db 181 LGVKTSNWLFEAGRAKLFNRLTGGVFISAGVMSTANRS 219

RESULT 13
Q92NK8 PRELIMINARY; PRT; 211 AA.
AC Q92NK8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Putative amino acid efflux protein.
GN R02191 OR SMC01851.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
```

